Functional, biochemical and structural analyses of *Plasmodium falciparum* pyruvate dehydrogenase complex

by

Marjo Larissa Laine BSc (Hons)

Thesis submitted for the degree of Doctor of Philosophy

September 2013

Institute of Infection, Immunity & Inflammation
College of Medical, Veterinary & Life Sciences
University of Glasgow

Abstract

The apicomplexan parasite *Plasmodium* is the causative agent of the devastating tropical disease, malaria. The World Health Organisation reported that in 2010 there were an estimated 219 million malaria cases and about 660,000 deaths. Sub-Saharan Africa is the worst affected, with 80% of malaria deaths and 90% of cases occurring in this area of the world. *P. falciparum* causes the most severe form of the illness and accounts for the majority of malaria deaths. Parasite resistance to antimalarials including the most effective drug, artemisinin, is becoming an increasing problem, thus research into new drug targets is vital.

The pyruvate dehydrogenase complex (PDC) is one of the α-ketoacid dehydrogenase complexes, which are involved in energy and amino acid metabolism. The PDC catalyses the transfer of the acetyl group from pyruvate to coenzyme A (CoA) to form acetyl-CoA. The complex comprises three enzymes; pyruvate dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and dihydrolipoamide dehydrogenase (E3). E2 has a multi-domain structure where the C-terminal catalytic domains (CD) of several E2 subunits interact to form the core of the complex. This core is an icosahedral 60-mer in mammals, plants and Gram-positive bacteria and an octahedral 24-mer in Gram-negative bacteria. The N-terminal part of E2 consists of the sub-unit binding domain (SBD) and 1 to 3 lipoyl-domains (LD) creating the 'swinging arm' of the PDC, which facilitates substrate channelling. E1 (a heterotetramer) and E3 (a homodimer) bind to the SBD to form the functional PDC.

In humans, PDC converts pyruvate to acetyl-CoA, which leads into the citric acid cycle located in the mitochondrion. *P. falciparum* PDC, however, is found only in the apicoplast and produces acetyl-CoA for fatty acid biosynthesis. *Plasmodium* PDC has recently been shown to be important for parasite progression from the asymptomatic liver stage to the symptomatic erythrocytic stage. Thus, inhibiting PDC could prevent development of malaria.

This study focuses on the role of the PDC in *P. falciparum* blood stages and the identification and characterisation of structural and biochemical differences between parasite and human PDC that may be ultimately exploited for drug or vaccine development.

I have optimised the expression and purification of soluble recombinant mature-length *P. falciparum* (*Pf*) E2 (His-r*Pf*E2m), truncated *Pf*E2 consisting of the SBD and CD (His-r*Pf*E2bc) and mature-length apicoplast *Pf*E3 (His-r*Pf*aE3) to obtain mg amounts of protein for biochemical and structural analyses. Each of the recombinant proteins was catalytically active.

Analytical ultracentrifugation (AUC) showed that His-r*Pf*E2m forms the typical trimer building blocks required for the large E2 core formation. Sedimentation velocity (SV) experiments showed a main species with a sedimentation coefficient of 6.6 ± 0.1 S, which corresponded to a species consistent with the *Pf*E2 trimer size observed in sedimentation equilibrium (SE) studies. However, no 24-mer or 60-mer core species were detected in SV experiments. SE analyses did show some larger molecular mass species (≥ 1.5 MDa), however, whether these represent a *Pf*E2 core complex was inconclusive due to interference by aggregated protein in the sample. My conclusion from these data is that *Pf*E2 may form a large core structure but that this is very unstable compared with E2 multimers from other organisms, where the core complex is readily formed and maintained. Similar results were obtained from SV and SE analyses of His-r*Pf*E2bc; only trimers are present.

Small angle X-ray scattering (SAXS) was used to further analyse the trimer structure. The solution structure obtained for His-r*Pf*E2bc revealed the linker between the SBD and the CD is extended and partially flexible. This conformation, which would allow SBD interaction with E1 and E3, is supported by cryo-electron microscopy images obtained by others for E2 from other organisms.

AUC on His-r*Pf*aE3 showed a main species with a sedimentation coefficient of 6.1 ± 0.1 S and SE analyses of the protein confirmed it to be a dimer as expected. These findings were further corroborated with the solution structure obtained using SAXS.

Deletion of the *PfE2* gene from *P. falciparum* was unsuccessful, as the correct gene locus was not targeted. This could be due to the gene being required for parasite survival in the blood stages. However, it could also be due to the difficulty of *P. falciparum* gene manipulation and hence, further attempts to delete *PfE2* using different approaches will be required. The *Pf*aE3 gene was successfully deleted, thus as shown in previous studies with murine malaria species, the gene is not essential in the human malaria parasite, *P. falciparum*. The *PfaE3* deletion mutants did not show increased susceptibility to oxidative

stressors compared with the wild type strain, suggesting that *Pf*aE3 does not play a role in defence against oxidative stress. However, the mutants were less susceptible to triclosan, an inhibitor of the fatty acid biosynthesis enzyme FabI. In addition, the mutant parasites maintained synchronicity over more than four replication cycles as opposed to WT parasites, which gradually lost synchronicity over this period of *in vitro* culture. Further work will need to be carried out to fully characterise the role of *Pf*aE3 in *P. falciparum.*

Author's Declaration

I, Marjo Larissa Laine, hereby declare that I am the sole author of this thesis and performed all the work, with the following exceptions:

Chapter 3:

- Mass spectrometry identification of protein bands shown in Fig. 3-13F was carried out by the Proteomics Unit, University of Glasgow.
- The His-r*Pf*E2bc construct was generated by Jamie Whitelaw.
- SDS-PAGE and western blot in Fig. 3-25A & B were generated by Jamie Whitelaw.

Acknowledgements

First and foremost I'd like to thank my supervisors Prof Sylke Müller and Dr. Olwyn Byron for giving me the opportunity to carry out my PhD in their laboratories and for their guidance, encouragement and inspiration throughout my PhD. Sylke, thanks for mentoring me with all the biochemistry and parasite work. Olwyn, thanks for persevering with me while I got my head around the biophysical side of my project. It's been a challenging few years, and without the immense support I received from both of you I could not have got to where I am now – submitting my thesis!

I am extremely grateful to the Wellcome Trust and also past and present directors of the Glasgow University Wellcome Trust 4-year PhD Programme, Prof Darren Monkton, Dr. Olwyn Byron and Prof Bill Cushley. Life does not always go to plan, but for giving me a chance regardless to pursue my ambition of becoming a research scientist I am forever indebted.

A great big thanks goes to all the current and previous members of the Müller group – Janet, Eva, Ellie, Andy, Tania, Anne, Sonal, and Marco, thanks for all the help with general lab stuff and all the laughs, lunches and drinks we've shared. To each of you, who at some point or another took care of my parasites, thanks! A special thank you goes to Ellie, for teaching me how to clone, putting up with all my questions and out of the blue flying off to Hamburg on a SAXS trip with me. Thanks for being not just a colleague but also a great friend and brightening up long days in the lab. A huge thanks to Jamie for being a model student and helping during his placement with me.

I'd also like to thank my assessors, Prof Gordon Lindsay and Dr. Nisha Philip for their invaluable advice throughout my project at meetings, poster sessions and talks. A thanks goes to the staff at the EMBL outstation at the Deutsches Elektronen Synchrotron in Hamburg, Germany, for their help during each of my trips out there. Thank you Mattia Rocco and Emre Brooks for your help with using US-SOMO.

An enormous thank you to my mum and dad, Liisa and Antti, for always believing in me and being a source of encouragement, advice and spiritual strength. You always said you could see me studying in university and how right you were! A big thanks to my sister Susanna and my lovely niece and nephew Tiia and Tomi, who have all made my holidays back home awesome by giving me something else to focus on and making me forget about the hard times. Carita, Marko, Christa and Kim, thanks for all the unconditional support and making me feel part of the family. Thank you Emilia for always being there for me even when it took me ages to respond to messages when I was all but sleeping in the lab. Your jokes, humour and kind words kept me going more than you know. Thanks Caro, Mark and Roope for being exactly who you are and reminding me to be exactly who I am too.

Last, but definitely not least, I owe everything to my husband, Marco, who has been my soul mate and a never-ending source of hugs. Also, a huge thank you for being my computer guru! I would not have survived the past few years without you and I am eternally grateful for your infinite patience and assurance through the tears, tantrums and tough times. Just five minutes with you and the furries (Snow, Blue, Roxy, Nikki, Iggie, Gregor, Hector, Teddy and Pookie) made a bad day better and a good day awesome. This is the ending to a long chapter in our lives but just the beginning of a new one. I can't wait to see what's in store for us next.

I tried and failed. I tried again and again and succeeded.

-Inscription on Gail Borden's gravestone.

Table of Contents

List of figures

List of Tables

Abbreviations

γGCS	Gamma-glutamylcysteine synthetase
5-FC	5-fluorocytosine
A_{260}	Absorbance at 260 nm
A_{280}	Absorbance at 280 nm
aa	Amino acids
AUC	Analytical ultracentrifugation
ACT	Artemisinin based combination therapies
Amp	Ampicillin
At	*Arabidopsis thaliana*
Av	*Azotobacter vinelandii*
Å	Ångstrom
bp	Base pairs
BSA	Bovine serum albumin
BSD	Blasticidin-S-deaminase
BSO	L-buthionine sulfoximine
CD	Catalytic domain
C-terminal	Carboxy terminal
CoA	Coenzyme A
CyD	Cytosine deaminase
CV	Column volume
Da	Dalton
dH_2O	Distilled water
DHLA	Dihydrolipoic acid
D_{max}	Maximum particle diameter
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
E1	Pyruvate dehydrogenase
E2	Dihydrolipoamide acetyltransferase
E3	Dihydrolipoamide dehydrogenase
E3BP	E3 binding protein
Ec	*Escherichia coli*
EDTA	Ethylene diamine tetraacetic acid
EM	Electron microscopy
FAD	Flavin adenine dinuceotide
fwd	Forward
gDNA	Genomic DNA
Gs	*Geobacillus stearothermophilus*
GSH	Glutathione (reduced form)
GST	Glutathione-S-transferase
hDHFR	Human dihydrofolate reductase
HEPES	4-(2-Hydroxyethyl)-piperazineethanesulphonic acid
Hu	*Homo sapiens*
IC_{50}	Concentration at which 50 % growth inhibition occurs
IPTG	Isopropyl β-D-thio-galactopyranoside
Kan	Kanamycin
kb	Kilobases
kDa	Kilodaltons
LB	Luria Bertani
LD	Lipoyl domain

MDa	Megadaltons
MES	2-(N-morpholino)ethanesulfonic acid
MOPS	3-(N-morpholino)propanesulfonic acid
MWCO	Molecular weight cutoff
NAD^+	Nicotinamide adenine dinucleotide (oxidized)
NADH	Nicotinamide adenine dinucleotide (reduced)
$NADP^+$	Nicotinamide adenine dinucleotide phosphate (oxidized)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
Ni-NTA	Nickel-nitriloacetic acid
NMR	Nuclear magnetic resonance
NSD	Normalised spatial discrepancy
N-terminal	Amino terminal
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
Pb	*Plasmodium berghei*
PBS	Phosphate buffered saline
Pc	*Plasmodium chabaudi*
PCR	Polymerase chain reaction
PDC	Pyruvate dehydroganse complex
Pf	*Plasmodium falciparum*
Pk	*Plasmodium knowlesi*
PMSF	Phenylmethanesulfonyl fluoride
Pv	*Plasmodium vivax*
Py	*Plasmodium yoelii*
rev	Reverse
R_g	Radius of gyration
rpm	Revolutions per minute
rmsd	Root mean square deviation
s	Seconds
s	Sedimentation coefficient
S	Svedberg units (10^{-13} s)
SAXS	Small angle X-ray scattering
SBD	Sub-unit binding domain
SDS	Sodium dodecyl sulphate
SE	Sedimentation equilibrium
SOC	Super optimal growth with catabolite repression
SV	Sedimentation velocity
Sy	*Synechocystis* sp. PCC 6803
TAE	Tris-acetate containing EDTA
TB	Terrific broth
TE	Tris containing EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
Tfb I	Transformation buffer 1
Tfb II	Transformation buffer 2
TPP	Thiamine diphosphate
Tris	Tris [hydroxymethyl] aminomethane
UV	Ultraviolet
V_e	Elution volume
V_0	Void volume
v/v	Volume per volume
w/v	Weight per volume
WHO	World Health Organisation
WT	Wild type

Chapter 1
Introduction

1.1 The *Plasmodium* parasite

The *Plasmodium* species are intracellular protozoan parasites of the phylum Apicomplexa, which also includes *Toxoplasma, Cryptosporidium,* and the important animal pathogens *Babesia, Theileria* and *Eimeria* (Feagin, 1994). These organisms possess an apical complex (hence the name Apicomplexa) comprising rhoptries, dense granules and micronemes, which are involved in parasite invasion into host cells and the formation of a parasitophorous vacuole. Another characteristic of Apicomplexa is a complex life cycle involving several hosts (Feagin, 1994).

There are over 100 species of *Plasmodium*, however, of these only five can infect humans (Tuteja, 2007). These are *Plasmodium falciparum,* the causative agent of the most severe form of malaria and the highest number of malaria cases and deaths worldwide (WHO, 2012). *P. vivax,* which is most widespread but seldom causes fatal illness (Mueller *et al.*, 2009) and the rarer *P. ovale* (Collins and Jeffery, 2005) and *P. malariae* (Collins and Jeffery, 2007). *P. knowlesi,* the simian parasite, has also been reported to infect humans in South-East Asia and has the ability to cause fatal disease (Cox-Singh *et al.*, 2010; Cox-Singh, 2012). Other species of interest are the bird parasites *P. lophurae* and *P. gallinaceum* and rodent malaria parasites *P. berghei, P. yoelii, P. chabaudi* and *P. vinckei,* which were used in research before *in vitro* culture of *P. falciparum* was developed in the 1970s (Trager and Jensen, 1976). Since then, the use of bird parasites has declined, however, the rodent parasites are still extensively used as an *in vivo* disease model for malaria, as culture of the exo-erythrocytic liver stages of *P. falciparum* is still extremely difficult (Sherman, 2009b). Also the rodent malaria parasites are more amenable to genetic manipulation making them valuable tools in studying *Plasmodium* biology (Limenitakis and Soldati-Favre, 2001).

Sequencing of the *P. falciparum* genome has taken research of the parasite into the postgenomic era. The genome is about 23 Mb, is organised into 14 chromosomes and has an unusually high A/T content of 80.6% (Gardner *et al.,* 2002a). In addition, the parasite has a circular 35 kb plastidial genome (Wilson *et al.,* 1996) and a 6 kb linear mitochondrial

genome (Sherman, 2009a). Around 5000 proteins are predicted to be encoded by the nuclear genome, 60% of which are hypothetical proteins without similarity to any known proteins (Gardner *et al.*, 2002a). Protein expression is highly regulated as the expression patterns change frequently while the parasite progresses through the different forms in its life cycle (Bozdech *et al.*, 2003).

1.1.1 The vectors

Around the turn of the 20th century Ronald Ross concluded that the mosquito was the vector for malaria. The conclusion was further developed by Giovanni Battista Grassi, who showed that it was the *Anopheles* mosquito that transmitted malaria (Sherman, 2005). There are 400 species of *Anopheles* mosquitoes of which ~60 are malaria vectors and of these about 30 are of importance to transmission (Tuteja, 2007). *Anopheles gambiae* is by far the most efficient malaria transmitter in sub-Saharan Africa (Collins and Hill, 2005). This is followed by *A. arabiensis*, which unlike *A. gambiae*, tolerates cooler and drier environments, thus extending malaria transmission further north and south and also into the drier seasons of the year (Collins and Hill, 2005).

1.1.2 The life cycle

The complex life cycle of *Plasmodium* (Figure 1-1) progresses from the sexual stages in female *Anopheles* mosquitoes to the asexual stages in the human host (Sherman, 2005). The two stages relevant to this study are the liver and intraerythrocytic stages.

1.1.2.1 Liver stage

Sporozoites residing in the mosquito salivary glands are injected into a human as the mosquito takes a blood meal. Typically about 20 sporozoites enter the human host and after migrating in the skin tissue for at least 30 min (Vanderberg and Frevert, 2004), they migrate towards the liver. Once in the sinusoid the sporozoites adhere to the endothelial cells, glide along the surface until they encounter hepatic macrophages called Kuppfer cells and traverse through these to get to hepatocytes (Frevert *et al.*, 2005).

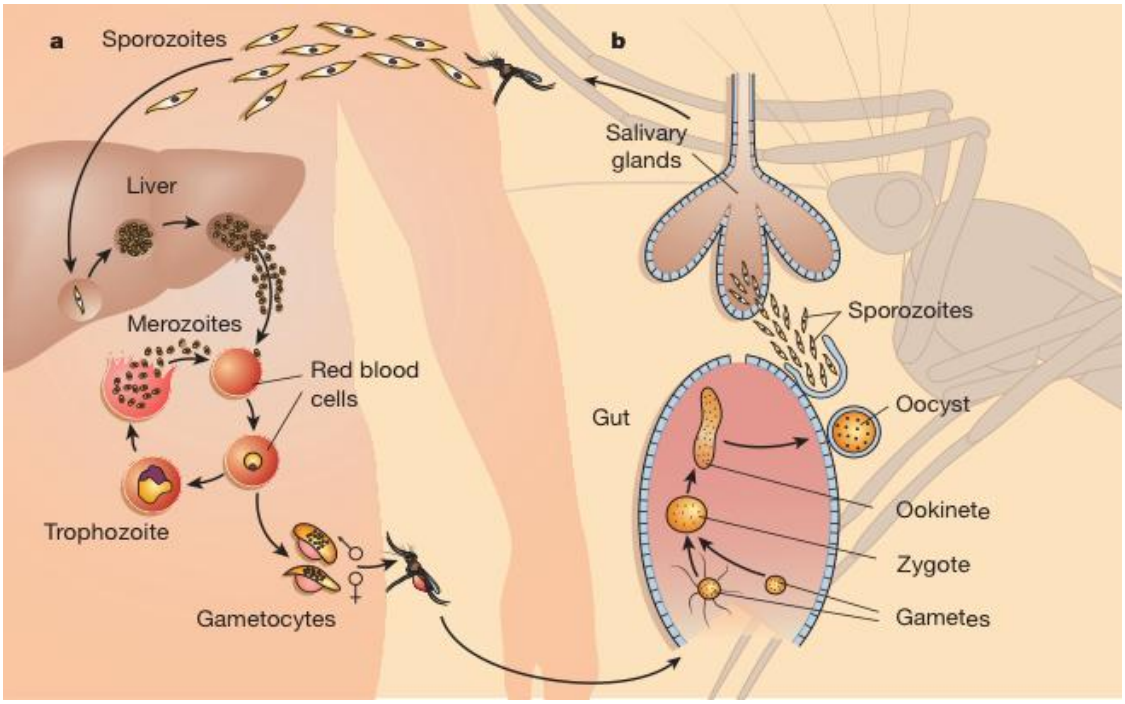


Figure 1-1. The malaria parasite life cycle.
(A) Asexual life cycles in the human host. (B) Sexual and asexual stages in the mosquito vector. Sporozoites are injected into a human as the mosquito takes a blood meal. The sporozoites then invade hepatocytes and multiply asexually to form merozoites, which go on to infect erythrocytes. In the red blood cells the parasites go through another cycle of asexual replication, which consists of the ring, trophozoite and schizont stage and the development of merozoites. The released merozoites go on to infect new erythrocytes. Some merozoites undergo gametocytogenesis, to produce male and female gametes, which are ingested by the female mosquito. In the mosquito gut, the female macrogamete is fertilised by a male microgamete to form an ookinete, which migrates through the mosquito gut wall to form an oocyst. Asexual replication takes place in the oocyst to produce sporozoites. The oocyst bursts to release sporozoites, which migrate into the mosquito salivary glands to wait for the mosquito to take its next blood meal (Sherman, 2005). This figure was taken from Wirth (2002), with permission from Nature Publishing Group.

For reasons not yet fully understood, the parasites traverse through several hepatocytes before invading one and initiating liver stage development (Mota *et al.*, 2001). On invasion, the sporozoite forms a surrounding parasitophorous vacuole (PV) using the host plasma membrane, which protects the parasite from the cytoplasm of the host cell (Mikolajczak and Kappe, 2006). This is followed by de-differentiation of the sporozoite into the liver trophozoite, which takes 5-7 days in the human parasite (Mikolajczak and Kappe, 2006). During the intermediate growth phase, the parasite prepares to exploit the host cell for the following rapid growth phase, or schizogony, during which extensive DNA and organelle replication occurs to form merozoites (Lindner *et al.*, 2012). In the late liver stages, the parasite grows larger than its host cell, the PV becomes damaged and degrades, which induces hepatocyte death and detachment from the surrounding tissue

(Sturm *et al.*, 2006; Sturm *et al.*, 2009). The merozoites are pushed into vacuoles called merosomes, which protrude from the hepatocyte into the sinusoids (Sturm *et al.*, 2006). The merosomes are then passively transported in the blood to the heart and lungs, where the merosomes rupture and the merozoites enter the blood stream to initiate the erythrocytic stages of the life cycle (Baer *et al.*, 2007).

1.1.2.2 Intraerythrocytic stage

Merozoites recognise surface receptors on erythrocytes to which the merozoite attaches and then re-orientates so that the apical end of the parasite is adjacent to the erythrocyte (Cowman and Crabb, 2006). A tight junction is formed and the parasite pushes into the erythrocyte while simultaneously shedding the surface coat of ligands required for invasion (Harris *et al.*, 2005). In the final stage the parasite invades the erythrocyte and forms the PV (Cowman and Crabb, 2006). Once inside the red blood cell, asexual replication begins. The merozoite becomes a discoidal ring stage parasite, which grows into a highly metabolically active trophozoite (Bannister *et al.*, 2000). The trophozoite grows and changes shape while consuming nutrients and haemoglobin from the host, resulting in extensive remodelling of the erythocyte (Goldberg and Cowman, 2010). The PV grows to be adjacent to the host plasma membrane and towards the end of the trophozoite stage (28-31 h after *P. falciparum* merozoite invasion) DNA replication is initiated (Arnot and Gull, 1998). During the schizont stage the nucleus divides approximately 4 times to produce, in most cases, 16 merozoites (Bannister *et al.*, 2000). The steps for egress of the merozoites from erythrocytes are initiated already soon after infection is established, as the parasite begins to destabilise the red blood cell membrane by incorporating plasmodial proteins into the submembraneous skeleton of the host cell (Wirth and Pradel, 2012). Several parasite proteases, including cysteine protease falcipain-2, aspartic protease plasmepsin II, subtilisin-like parasite protease *Pf*SUB1, as well as the host protease erythrocyte-derived calpain-1, have been shown to be involved in parasite egress (Wirth and Pradel, 2012). *Pf*SUB1 was shown to be released into the PV space from a secretory organelle called an exoneme (Yeoh *et al.*, 2007), thus it is thought that the parasite proteases involved in egress would be contained in the exoneme until required (Wirth and Pradel, 2012). Following a yet unidentified signal for merozoite egress, the proteases are released and degrade the PV, which results in an intake of water from the erythrocyte leading to the rupture of the PV by an unknown mechanism (Glushakova *et al.*, 2010; Wirth and Pradel,

2012). Subsequently a protease independent pore formation occurs in the erythrocyte membrane which then curls outwards and turns inside out to propel the merozoites into the blood stream (Glushakova *et al.*, 2010; Abkarian *et al.*, 2011). Once in the bloodstream, the merozoites are short lived and take only about 60 seconds to invade the next host erythrocyte (Cowman and Crabb, 2006). In some cases, merozoites do not continue this cycle but undergo gametocytogenesis to produce female and male gametes, which are then ingested by the female mosquito as it takes a blood meal (Baker, 2010).

The intraerythrocytic stage of the infection in humans varies in the time it takes to undergo intraerythrocytic schizogony. *P. falciparum, P. vivax* and *P. ovale* undergo the cycle in 48 hours, while it takes 72 hours for *P. malariae* and 24 hours for *P. knowlesi*. It is this developmental stage that causes the symptoms of the disease and the release of merozoites at the end of the erythrocytic cycle coincides with fever, thus the names tertian, quartan and quotidian malaria, respectively (Grobusch and Kremsner, 2005).

1.2 Malaria

The name malaria comes from the assumption that malaria was spread by 'mal' aria', or bad air from the swampy land that surrounded areas where the illness occurred (Tuteja, 2007). Records of 'autumnal fevers' in texts from South Asia and China dating back to 1500 to 800 BCE, give some clue to the antiquity of the disease (Packard, 2007). The disease, originating from Africa, also moved north into Europe. While *P. falciparum* did not reproduce well in the northern climate, *P. vivax* and *P. malariae* did, even as far as northern Russia. In the 17th century, in south-eastern England malaria was a significant source of morbidity and mortality and the disease had also made its way to the "New World", as early settlers migrated to America (Packard, 2007). The arrival of *P. falciparum* malaria onto the continent is thought to be coincident with the arrival of slaves brought from Africa (Packard, 2007).

Today, malaria is a tropical disease. Europe and North America were able to eradicate the illness following an increase in the socio-economic status of these countries pushed forward by development of agricultural methods, improved hygiene, better housing to protect from the mosquito vectors and access to the antimalarial quinine (Packard, 2007).

As these developments have lagged behind in Africa, Asia and South America, the illness has become almost solely a tropical blight (Packard, 2007).

General symptoms of malaria are the periodic fever stated above, as well as headache, fatigue and malaise, musculoskeletal pain, nausea, vomiting and in some cases diarrhoea (Grobusch and Kremsner, 2005). *Falciparum* malaria accounts for 93% of malaria cases (Fig. 1-2) as well as the majority of deaths worldwide (WHO, 2012). *P. falciparum* is the deadliest of the malaria parasites as it is most likely to cause severe disease. This is characterised by malarial anaemia, acidosis and the development of cerebral malaria, which are the three primary causes of death in patients (Grobusch and Kremsner, 2005).

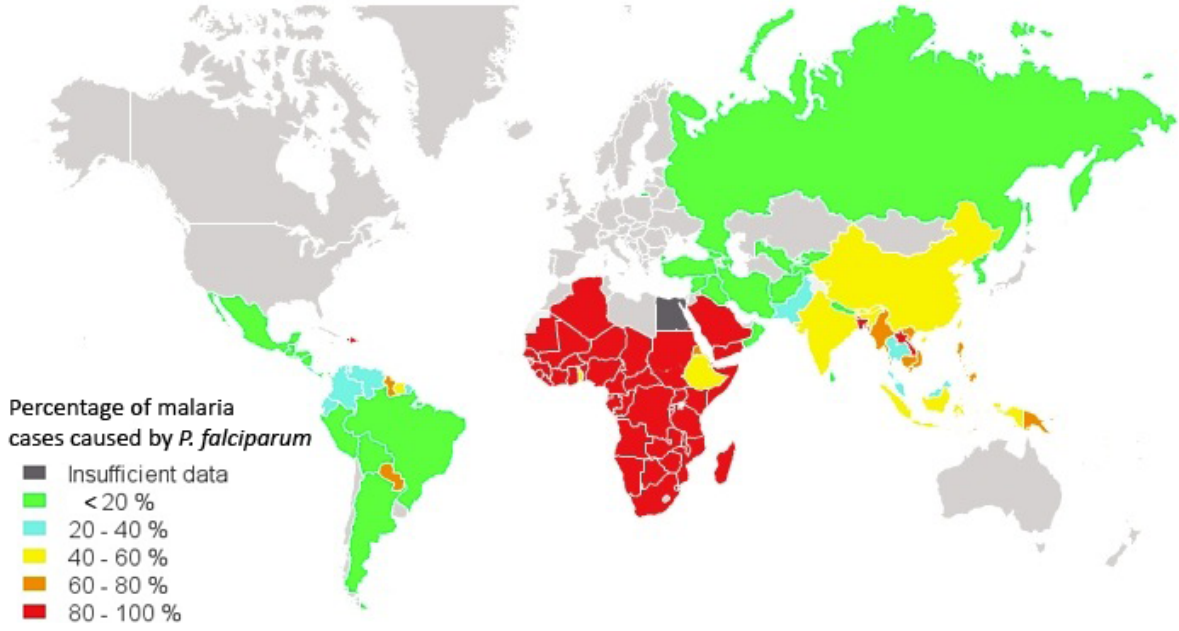


Figure 1-2. Global distribution of *P. falciparum* malaria cases.
Countries coloured in light grey do not have malaria. Figure taken from the Global Malaria Mapper created by Medicines for Malaria Venture and the WHO Global Malaria Programme from data obtained from the WHO World Malaria Report, 2012 (http://www.worldmalariareport.org/). Taken with permission from the World Health Organisation.

The World Health Organisation (WHO, 2012) reports that globally in 2010 there were an estimated 219 million malaria cases and about 660,000 deaths. Sub-Saharan Africa is by far the worst affected with 80% of cases and 90% of deaths in this region. Children under the age of five are especially at risk as 85% of the deaths occur in this age group. Malaria accounts for 20% of all childhood deaths. Pregnant women are also at high risk of infection, which can result in complications in the pregnancy and even death of the mother and child. Half the world's population is at risk of acquiring the illness. Malaria's gravity does not end at loss of life, but the economic loss due to periods of illness and post-infection complications is estimated to be £8.3 billion a year (WHO, 2009). This is a substantial loss to the affected countries as their populations are already generally poor. As poverty and malaria tend to exacerbate each other, eradication of the illness will require not only efficient antimalarials to combat the symptoms and prevent deaths but other control measures and socio-economical development of these countries (Packard, 2007).

1.2.1 Treatment and disease control

Cinchona bark was effectively used as an antimalarial in ancient times (Packard, 2007). Quinine, the active agent in the bark, was later used to treat malaria and was widely used in Britain and North America in the eradication of the disease. The quinine derivative, chloroquine, was very effective and cheap, thus widely used, however, parasite resistance to the drug has made it largely ineffective especially in Africa and Asia (Fig. 1-3). Other antimalarial drugs include other quinine derivatives such as mefloquine and primaquine, the antifolate drug sulfadoxine-pyrimethamine and the naphthalene derivative atovaquone (White, 2008). However, resistance to these is also increasing (WHO, 2010).

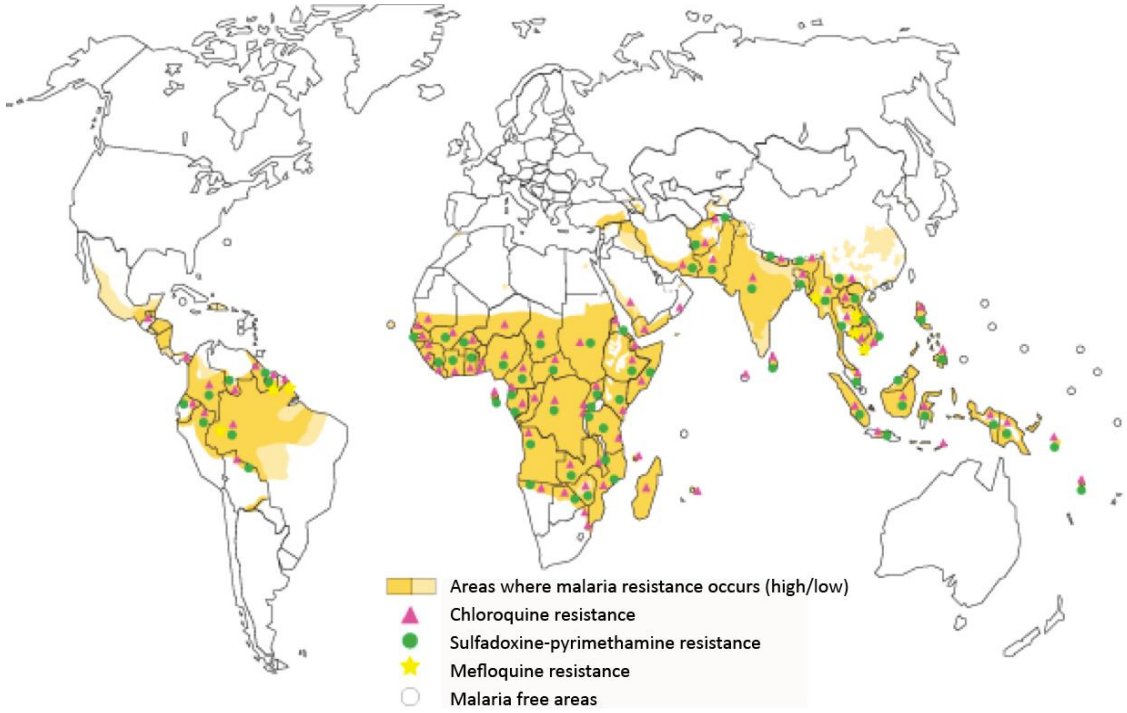


Figure 1-3. *P. falciparum* resistance to the commonly used antimalarials (WHO, 2005). Taken from the World Malaria Report (2005) with permission from the World Health Organisation.

Drugs based on artemisinin, which is derived from the plant *Artemisia annua*, remain extremely effective, although parasite resistance has been reported in some countries in South-East Asia (WHO, 2012). To reduce the occurrence of resistance, artemisinin combination therapy (ACT) is the recommended first-line therapy. Unfortunately, most of those who would require ACT do not receive it due to its high cost (White, 2008). For example, a course of sulfadoxine-pyrimethamine in Kenya costs $0.38 whereas a course of Coartem ACT treatment is $7.60 (Packard, 2007). No vaccine against malaria is yet available.

The increase in resistance to antimalarials creates an urgent need for research to identify and validate new drug targets, so that combination therapy can be used for future treatment and prophylaxis. Drugs alone will not eliminate malaria, and it is likely that eradication of the disease will require a combined effort of effective drug treatment, vector control methods and socio-economic development of the endemic countries. However, as White (2008) reviews, if we lose the current effective drugs, which at the moment are mainly just the artemisinin based drugs, the successful control and elimination of malaria will become impossible.

1.3 The *Plasmodium* apicoplast

The apicoplast was already seen on microscope images of avian *Plasmodium* parasites in the 1960's. At the time it was called a 'spherical body' and it was shown to be closely associated with mitochondria (Aikawa, 1966; Hepler *et al.*, 1966, Fig. 1-4). A circular 10.3 μm DNA was shown in electron micrographs following sub-cellular fractionation of *P. lophurae* and was presumed to be mitochondrial (Kilejian, 1975). Subsequent studies showed the presence of the circular ~10-12 μm and ~ 35 kb DNA in a wide range of apicomplexan parasites including *P. falciparum*, however, for decades the genome was incorrectly thought to be mitochondrial (Dore *et al.*, 1983; Borst *et al.*, 1984; Williamson *et al.*, 1985; Gardner *et al.*, 1988). A second linear 6kb DNA was also discovered, which was later shown to be the real mitochondrial genome (Vaidya and Arasu, 1987; Vaidya *et al.*, 1989; Feagin *et al.*, 1991). It was not until the 1990's when the enigma of the circular 35 kb DNA in the parasite was solved. Inconsistencies of the genome with those of

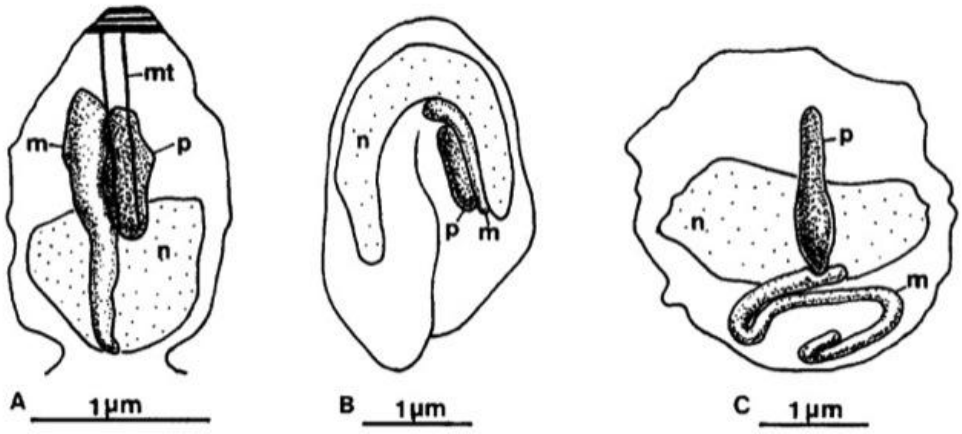


Figure 1-4. The close association of the mitochondrion and apicoplast shown in different intra-erythrocytic stages of *P. falciparum*.
(A) budding merozoite, (B) young ring, (C) young trophozoite. p, apicoplast; m) mitochondrion; mt) subpellicular microtubules; n) nucleus. Figure taken from Hopkins *et al*. (1999) with permission from Elsevier.

mitochondrial origin prompted suggestions of a chloroplast ancestry (Gardner *et al.*, 1991; Wilson *et al.*, 1991). Wilson *et al*. (1992) showed the 6 kb and 35 kb DNAs to localise to different sub-cellular fractions. This was followed by studies showing the 35 kb DNA to have features in common with plastid genomes (Williamson *et al.*, 1994). Finally, in 1996,

Wilson and colleagues published the complete gene map of the *P. falciparum* apicoplast DNA, which opened research into the exciting new organelle.

The general consensus today is that the apicoplast is of secondary endosymbiont origin, where the parasite has engulfed another eukaryote, most likely a single-celled red algae, which was a primary endosymbiont that obtained a plastid from a cyanobacterium (Botte *et al.*, 2011; van Dooren and Striepen, 2013). Because of the prokaryotic origin of the apicoplast, many of the enzymes and metabolic pathways in the organelle are bacterial, thus making it a pool of potential novel drug targets (Qidwai and Khan, 2012; Goodman and McFadden, 2013).

Apart from *Cryptosporidium*, all apicomplexan parasites studied so far contain an apicoplast (Ralph *et al.*, 2004). The organelle has four membranes, associates closely with the mitochondrion throughout the parasite's life-cycle (Hopkins *et al.*, 1999) and is essential for parasite survival (Fichera and Roos, 1997; Sullivan *et al.*, 2000). The 35 kb genome has a high A/T content of 86.9% and encodes mainly housekeeping proteins involved with translation and transcription of genes (Wilson *et al.*, 1996). In addition, nuclear encoded proteins are targeted to the apicoplast by N-terminal leader sequences. Each bipartite leader sequence consists of a hydrophobic 16-34 amino acid signal peptide (SP) and a 30-42 amino acid transit peptide (TP) with a positive charge and a high lysine and asparagine content (Waller *et al.*, 1998). The signal peptide targets the proteins to the endomembrane system, namely the endoplasmic reticulum, where it is cleaved off to expose the TP (Waller *et al.*, 2000). This targets the protein to the apicoplast, via a poorly understood mechanism where vesicles containing the proteins bud from the ER and fuse with the outer membrane of the apicoplast (Deponte *et al.*, 2012). Following transport across the apicoplast membranes, the TP is cleaved off to reveal the functional mature protein (Deponte *et al.*, 2012). Bioinformatic tools, which detect these leader sequences, have been used to predict that over 540 nuclear encoded proteins are targeted to the apicoplast (Ralph *et al.*, 2004).

1.3.1 Apicoplast metabolic pathways

Several metabolic pathways have been shown to localise to the apicoplast, including type II fatty acid biosynthesis (FAS II) (Waller *et al.*, 1998), non-mevalonate isoprenoid biosynthesis, synthesis of iron-sulphur clusters (Ralph *et al.*, 2004), lipoic acid biosynthesis (Wrenger and Müller, 2004) and part of the haem biosynthesis pathway (Sato *et al.*, 2004; Okada, 2009). The FAS II pathway and the lipoic acid biosynthesis pathways are of most relevance to this study and will be further discussed.

1.3.1.1 The fatty acid biosynthetic pathway

Until the turn of the 21st century it was thought that *Plasmodium* lacked the ability to synthesise fatty acids *de novo* and acquired the molecules solely by scavenging from the host red blood cell (Rock, 1971; Rock *et al.*, 1971; Holz, 1977; Holz *et al.*, 1977; Vial and Ancelin, 1992). However, Waller *et al*. (2000) showed that the parasite possessed nuclear encoded fatty-acid biosynthesis enzymes that were targeted to the recently discovered apicoplast. These enzymes included the acyl carrier protein (ACP), β-ketoacyl-ACP synthase III (FabH) and β-hydroxyl-ACP dehydratase (FabZ). This pathway requires acetyl-coenzyme A (acetyl-CoA) and/or malonyl-CoA, which enter a cycle of several consecutive elongations of acyl chains (Waller *et al.*, 2003, Fig. 1-5).

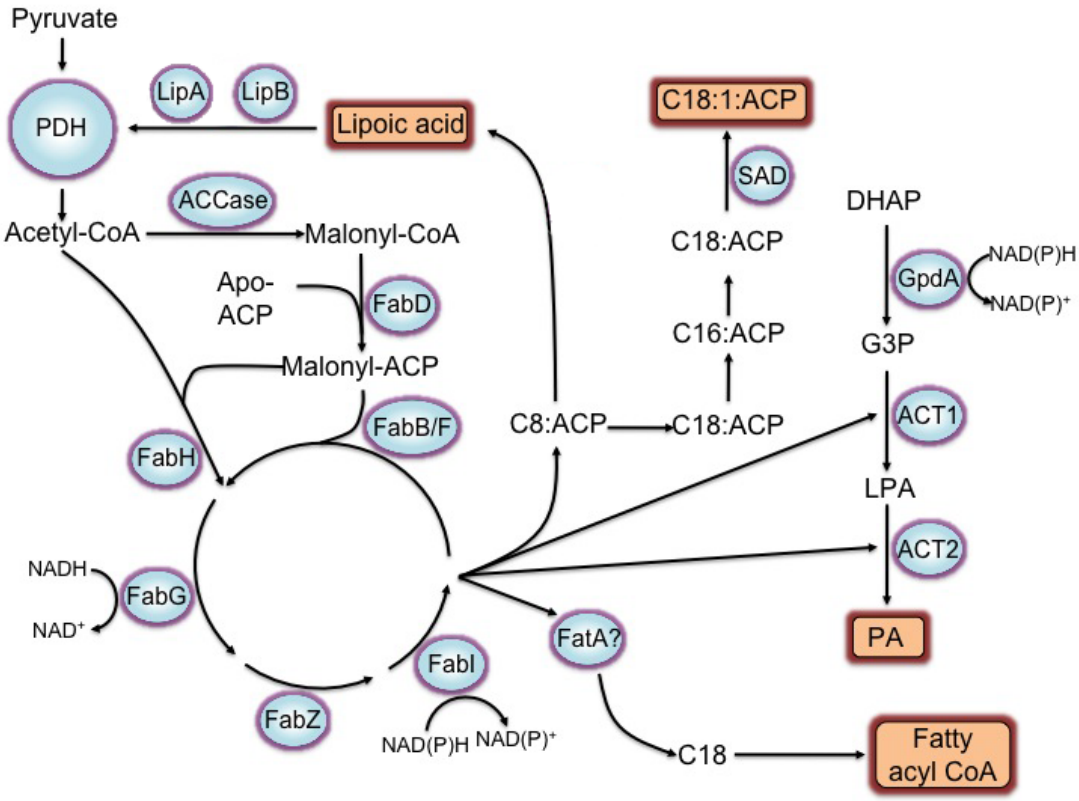


Figure 1-5. *P. falciparum* FAS II pathway.
Acetyl-CoA and/or malonyl-CoA feed into the several cycles of acyl chain elongation catalysed by FabH, FabG, FabZ, FabI and FabB/F. Products of FAS II include lipoic acid, free fatty acids such as oleic acid, phosphatidic acid and fatty acyl CoA. Substrates and intermediates in black, final products in brown boxes, enzymes in blue circles. ACCase, acetyl-CoA carboxylase; acetyl CoA, acetyl coenzyme A; ACP, acyl carrier protein; ACT1, glycerol-3-phosphate acyltransferase; ACT2, 1-acyl-glycerol-3-phosphate acyltransferase; FabB/F, β-ketoacyl ACP synthase I/II; FabD, malonyl-CoA transacylase; FabG, β-ketoacyl ACP reductase; FabH, β-keto-ACP synthase III; FabI, enoyl-ACP reductase; FabZ, β-hydroxyacyl-ACP dehydratase; FatA, acyl-ACP thioesterase; G3P, glyceraldehyde 3 phosphate; GpdA, glycerol-3-phosphate dehydrogenase; LipA, lipoic acid synthase; LipB, lipoate protein ligase; LPA, lysophosphatidic acid; NAD^+/NADH, nicotinamide adenosine; PA, phosphatidic acid; PDC, pyruvate dehydrogenase complex; SAD, stearoyl-ACP desaturase; C8, octanoic acid; C18, strearic acid; C16, palmitic acid; C18:1, oleic acid. Diagram adapted from Ralph *et al*. (2004).

The FAS II pathway is catalysed by seven separate enzymes, whereas the type I FAS pathway found in animals, including humans, posesses a large multifunctional fatty acid synthase complex (Waller *et al.*, 2003). The biochemical confirmation of the apicoplast FAS II was achieved by Surolia and Surolia (2001) as they showed ^{14}C labelled acetate- and malonyl-CoA incorporation into fatty acids both *in vivo* and *in vitro*. Additionally, bioinformatic analyses of apicoplast proteins have shown that all FAS enzymes exist in *Plasmodium* (Waller *et al.*, 1998; Waller *et al.*, 2000; McLeod *et al.*, 2001; Surolia and Surolia, 2001).

Fatty acids are required as precursors for several functions in the parasite such as biological membrane biogenesis and production of protein membrane anchors (Vaughan *et al.*, 2009). In addition, free fatty acids may be synthesised and transported out of the apicoplast possibly into the endoplasmic reticulum (Ralph *et al.*, 2004). Importantly, the FAS II pathway is essential for production of the pyruvate dehydrogenase complex (PDC) co-factor lipoic acid. The octanoyl-group from octanoyl-ACP, produced by the FASII pathway, is transferred by *N*-octanoyltransferase (LipB) to the lipoyl-domain of E2. This is followed by the addition of sulphurs to positions 6 and 8 of the octanoyl moiety by lipoic acid synthase (LipA) to form the fully functioning PDC (Storm and Müller, 2012).

As the *P. falciparum* mitochondrion does not synthesise the lipoic acid required for α-keto acid dehydrogenase complexes, but scavenges it from outside the organelle, it was suggested that lipoic acid is transported from the apicoplast to the mitochondrion. This would free the parasite from relying on a host supply of the co-factor (Wrenger and Müller, 2004). However, this hypothesis was disregarded after Crawford *et al*. (2006), showed that inhibition of apicoplast fatty acid biosynthesis in *Toxoplasma gondii* did not affect lipoylation of the mitochondrial proteins. Allary *et al*. (2007) and Günther *et al*. (2007) also showed that the mitochondrion had its own specific lipoic acid scavenging mechanism independent of the apicoplast *de novo* lipoic acid synthesis.

It was thought for almost a decade that FAS II was essential for *P. falciparum* during its intra-erythrocytic cycle as studies using FAS II pathway inhibitors inhibited parasite growth *in vitro* (Surolia and Surolia, 2001; Waller *et al.*, 2003). However, several groups showed that the enzymes involved in fatty acid biosynthesis are dispensable during intraerythrocytic development of rodent malaria species (Yu *et al.*, 2008a; Vaughan *et al.*, 2009). Nevertheless, it is has been shown that FAS II is essential for the development of the liver stages, since the majority of FAS II deficient parasites failed to form exo-erythrocytic merozoites and thus could not initiate the blood stage infection. This is discussed in more detail in section 1.4.7.1.

1.4 α-Keto acid dehydrogenase complexes (KADCs)

KADCs are a group of multienzyme complexes that are usually found in the mitochondrion and are involved in energy and amino acid metabolism by catalysing the irreversible oxidative decarboxylation of α-keto acids. This group consists of the pyruvate dehydrogenase complex (PDC), the α-ketoglutarate dehydrogenase complex (KGDC) and the branced-chain α-keto acid dehydrogenase complex (BCKDC). These are composed of three enzymes (E1, E2, E3) that catalyse similar reactions and function in close proximity to each other in a highly co-ordinated way to ensure correct substrate channelling (de Kok *et al.*, 1998; Mooney *et al.*, 2002).

Research into these large complexes started in the 1940 and 1950s with the isolation and characterisation of the 'acetate replacing factor' or 'pyruvate oxidation factor' i.e. α-lipoic acid, which is one of the essential co-factors of KADCs (Reed, 2001). Studies continued with the isolation and characterisation of PDC, KGDC and the BCKDC in the 1950s-1970s mainly by isolating the complexes from bacteria and yeast. The availability of molecular genetics methods from the 1980s onwards made it possible to clone, mutate and overexpress these proteins. This enabled research to be taken to the next level to acquire a better understanding of these carefully co-ordinated enzyme complexes.

1.4.1 The α-ketoglutarate dehydrogenase complex

The KGDC is located in the mitochondrion and is part of the tricarboxylic acid (TCA) cycle where it catalyses the irreversible decarboxylation of α-ketoglutarate to yield CO_2, succinyl-CoA and NADH (Tretter and Adam-Vizi, 2005). It is highly regulated and is the rate-limiting enzyme of the TCA cycle. The overall reaction is

$$\alpha\text{-ketoglutarate} + \text{CoA-SH} + \text{NAD}^+ \rightarrow \text{succinyl-CoA} + \text{CO}_2 + \text{NADH} + \text{H}^+ \qquad (1.1)$$

The kE1 enzyme, α-ketoglutarate decarboxylase, catalyses the oxidative decarboxylation of α-ketoglutarate by transferring the succinyl-group to the thiamine pyrophosphate (TPP) cofactor, releasing CO_2. The succinyl moiety is subsequently transferred onto the lipoic acid co-factor of kE2, the dihydrolipoamide succinyltransferase (kE2). This is followed by

the kE2 transfer of the succinyl-group to CoA to form succinyl-CoA, which is then used for the subsequent reactions in the TCA cycle or is used for heme biosynthesis (van Dooren *et al.*, 2006). The E3 is not specific to the KGDC but is also used by the other KADCs in the mitochondrion as well.

The mammalian KGDC has a molecular mass of 2 MDa, which is similar to that of the plant complex, which is 1.7 MDa. *P. falciparum* homologues of each of the KGDC enzymes exist and are likely to be targeted to the mitochondrion. kE2 and the E3 component have been shown experimentally to be targeted to the mitochondrion (McMillan *et al.*, 2005). The kE1 is a homodimer (Tretter and Adam-Vizi, 2005), while the bacterial kE2 is an octahedral 24-mer (Knapp *et al.*, 1998). The *P. falciparum kE1* derived amino acid sequence has 39% and *kE2* derived amino acid sequence 42% identity to the human gene (McMillan *et al.*, 2005).

KGDC is regulated by negative feedback inhibition by its products succinyl-CoA and NADH (Smith *et al.*, 1974) and not by reversible phosphorylation of the E1 enzyme as with the PDC and BCKDC. It has also been shown that Ca^{2+} levels influence KGDC activity by lowering the apparent K_m of kE1 for α-ketoglutarate (McCormack and Denton, 1979).

KGDC is easily targeted by reactive oxygen species (ROS), potentially due to the large amounts of suphydryl groups and tyrosines required for its activity (Starkov, 2013). Thus, ROS can cause decreased activity of the enzyme, which has been observed in neurodegenerative diseases such as Alzheimer's, Parkinson's and Huntington's disease (Starkov, 2013).

1.4.2 The branced-chain α-keto acid dehydrogenase complex

The BCKDC is involved in the catabolism of the amino acids isoleucine, leucine and valine (Brosnan and Brosnan, 2006). These amino acids are the most hydrophobic of the amino acids, thus playing a crucial role in globular and membrane proteins as well as in coiled-coil α-helices found in proteins such as myosin, keratin, fibrinogen and some transcription factors (Brosnan and Brosnan, 2006). A defect in the BCKDC interrupts the

catabolism of these amino acids resulting in their accumulation, which leads to maple syrup urine disease, a severe genetic disorder that can be fatal (Aevarsson *et al.*, 2000). The first step of branched-chain amino acid catabolism is catalysed by the branched-chain amino acid transaminase (BCAT), which converts isoleucine, valine and leucine to the respective α-keto acids α-keto-β-methylvalerate, α-ketoisovalerate and α-ketoisocaproic acid (Brosnan and Brosnan, 2006). The BCKDC catalyses the second step, where the aforementioned α-keto acid substrates are converted to acyl-CoA products α-methylbutyryl-CoA, isobutyryl-CoA and isovalyryl-CoA respectively. These products can then be converted to acetyl-CoA and/or succinyl-CoA for use in the TCA cycle (Ward *et al.*, 1999; Holecek, 2002). The overall reaction is as follows

$$\text{branched chain } \alpha\text{-keto acid} + \text{CoA-SH} + \text{NAD}^+ \rightarrow$$
$$\text{succinyl/acetyl-CoA} + \text{CO}_2 + \text{NADH} + \text{H}^+ \qquad (1.2)$$

As with KGDC, the BCKDC consists of three enzymes, homologues of which have been shown in *P. falciparum* and are likely to be localised to the mitochondrion (McMillan *et al.*, 2005). The branched chain α-keto acid decarboxylase (bE1) is composed of two α- and β-subunit dimers forming a heterotetramer. Regulation of the complex by phosphorylation and dephosphorylation is carried out by a branched chain α-keto acid dehydrogenase kinase and phosphatase at conserved serine residues in the bE1 enzyme (Odessey, 1982).

The *P. falciparum* bE1α and bE1β have 49% and 61% amino acid sequence identity to the human protein, whereas the *bE2* gene showed 34% identity (McMillan *et al.*, 2005). The dihydrolipoamide acetyltransferase (bE2) in mammals forms an octahedral 24-mer, and the E3 is again the same one shared with the other KADCs. The *P. falciparum* bE2 has been shown to be lipoylated during the intraerythrocytic stages, suggesting bE2 and possibly the whole complex is active (Allary *et al.*, 2007; Günther *et al.*, 2007). The role of the BCKDC in *P. falciparum* is unclear as the parasite lacks the gene encoding BCAT, thus the complex is not likely to be involved in the degradation of branch-chained amino acids (Seeber *et al.*, 2008; Storm and Müller, 2012). The BCKDC has been shown to be able to also catabolise pyruvate to acetyl-CoA (Pettit *et al.*, 1978; Heath *et al.*, 2007) thus it has been proposed that the *P. falciparum* BCKDC in the absence of a PDC could be the source of acetyl-CoA in the mitochondrion (Seeber *et al.*, 2008).

1.4.3 The pyruvate dehydrogenase complex

The PDC catalyses the conversion of pyruvate, the final end product of glycolysis, to acetyl-CoA which then enters the TCA cycle (Fig. 1-6) (de Kok *et al.*, 1998). The equation for this reaction is

$$\text{pyruvate} + \text{CoA-SH} + \text{NAD}^+ \rightarrow \text{acetyl-CoA} + \text{CO}_2 + \text{NADH} + \text{H}^+ \qquad (1.3)$$

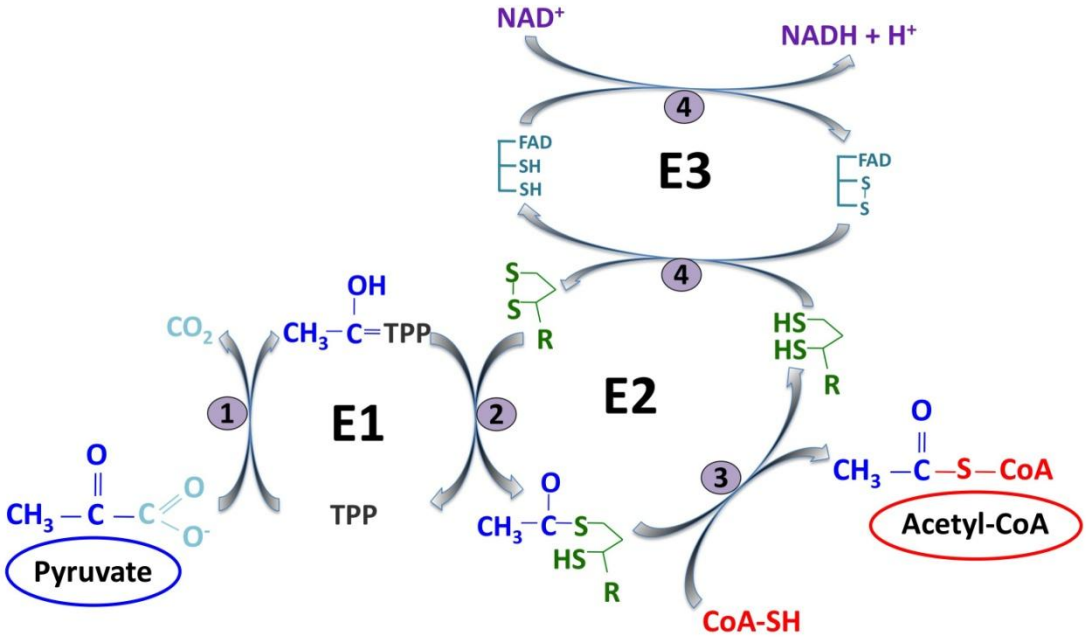


Figure 1-6. The reactions catalysed by the PDC.
(1) Oxidative decarboxylation of pyruvate by E1, (2) transfer of acetyl group from thiamine pyrophosphate (TPP) to lipoamide, (3) Acetylation of coenzyme A (CoA) by E2, (4) reoxidation of lipoamide by E3, (5) final electron transfer to NAD^+ yielding NADH + H^+. Diagram adapted from Mattevi *et al*. (1992a).

Regulation of the PDC in mammals is by reversible phosphorylation by pyruvate dehydrogenase kinase (PDK) and phosphatase (PDP). In addition, there is negative feedback inhibition by the products acetyl-CoA and NADH, while the presence of the substrates CoA, NAD^+, pyruvate as well as insulin and Ca^{2+} activate the PDC (Patel and Korotchkina, 2001). There are no PDK or PDP homologues in *P. falciparum*, thus regulation is likely to be by product inhibition, as in plants and bacteria where the

NAD^+/NADH and acetyl-CoA/CoA-SH ratios regulate the activity of the complex (de Kok *et al.*, 1998; Mooney *et al.*, 2002).

Genetic defects of the E1α subunit are most common, although mutations may occur in each of the enzymes of the PDC as well as the eukaryotic E3 binding protein (E3BP) (Patel *et al.*, 2012). Common symptoms of PDC deficiency include lactic acidosis and neurological and neuromuscular degeneration (Patel *et al.*, 2012). Another disease associated with the PDC is primary biliary chirrosis, an autoimmune liver disease, which is caused by anti-mitochondrial antibodies targeting mainly the lipoylated inner lipoyl domain of E2 (Hirschfield and Gershwin, 2013).

1.4.3.1 Pyruvate decarboxylase (E1)

The E1 of eukaryotes and Gram-positive bacteria is a heterotetramer composed of two subunits, E1α and E1β (Fig. 1-7), whereas in Gram-negative bacteria, the enzyme is a homodimer (Pei *et al.*, 2008).

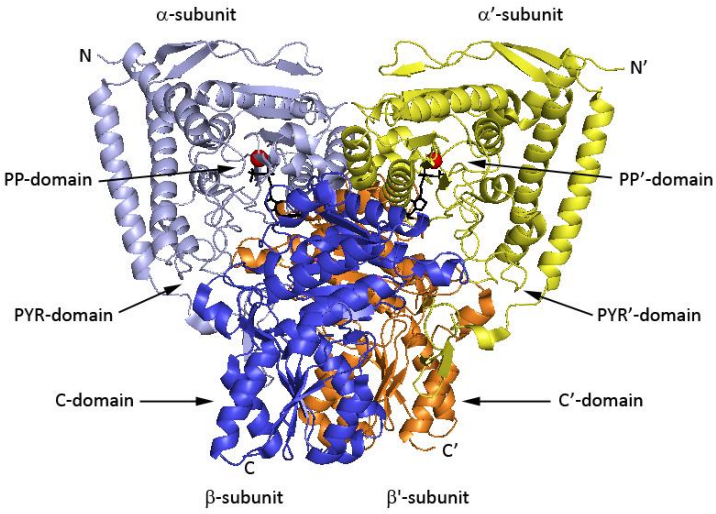


Figure 1-7. Human E1 structure
The heterotetramer E1 contains several domains. The PP-domain and PYR-domain are involved in tetramer and active site formation, while the C-domain binds the sub-unit binding domain (SBD) of E2 (Ciszak *et al.*, 2003; PDB entry 1NI4). Mg^{2+} shown as red spheres. Figure adapted from Ciszak *et al.*, 2003.

The E1α subunit contains what is termed the PP-domain, which is involved in binding the pyrophosphate of TPP and Mg^{2+}, another cofactor of the protein (Ciszak *et al.*, 2003). The E1β subunit forms two domains; the PYR-domain, which binds the aminopyrimidine ring of TPP, and the C-domain, which is involved in interactions with E2 (Ciszak *et al.*, 2003; Frank *et al.*, 2005). The $\alpha_2\beta_2$-heterotetramer is formed predominantly by hydrophobic interactions between the PP and PYR domains as well as more flexible interactions due to shared binding of TPP (Ciszak *et al.*, 2003).

E1 transfers the acetyl-group from the α-keto acid to TPP, concurrently releasing CO_2 (Fig. 1-8A). This enzyme confers substrate specificity to the PDC. The reaction catalysed by E1 takes place in two sequential reactions, thus it displays "ping-pong" kinetics (Pei *et al.*, 2008). In the first reaction, the acetyl group from pyruvate is transferred to TPP via a decarboxylation reaction to form the intermediate enamine-TPP. Subsequently, reductive acetylation takes place where the acetyl group is transferred onto one of the sulphurs on the lipoamide moiety of E2 to form S-acetyldihydrolipoamide-E2, while the other sulphur is reduced.

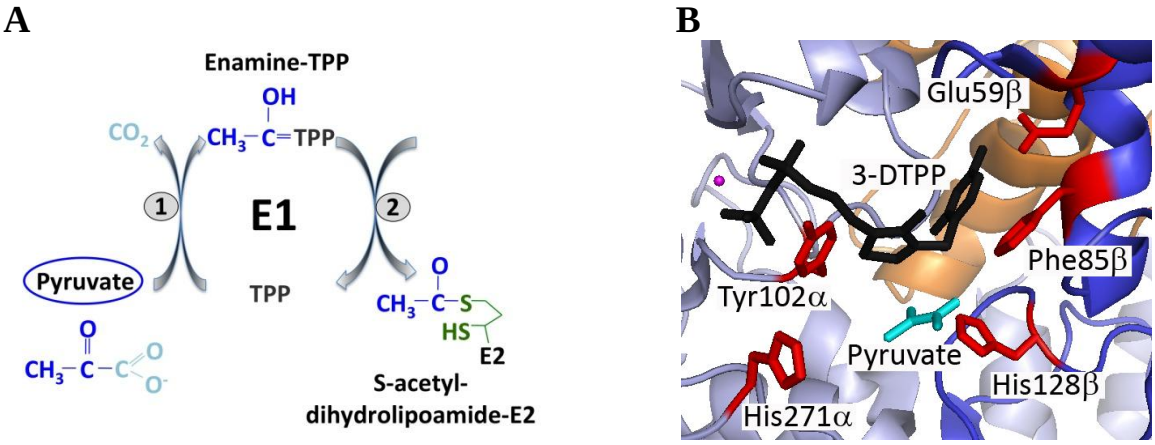


Figure 1-8. E1 reaction and active site
(A) The E1 reaction takes place in two steps; (1) the oxidative decarboxylation of pyruvate to form enamine-TPP and (2) the reductive transfer of the acetyl group to the lipoamide moiety of E2. (B) The active site of *G. stearothermophilus* E1 bound with the enamine-TPP analogue, 3-deazaTPP (3-DTPP, in black), pyruvate (cyan) and Mg^{2+} (red sphere). Residues in red are involved in the active site of E1, see text for description. Figure adapted from Pei *et al*., 2008.(PDB entry 3DV0).

Each E1 protein has two active sites, which are formed at the interface of the PP and PYR domains (Ciszak *et al.*, 2003; Pei *et al.*, 2008). Important residues in the active site include Glu59 (numbering based on the *Geobacillus stearothermophilus* E1), which initiates the proton transfer required to activate TPP, Phe85β and Tyr102α are important in enamine-TPP binding, His128β interacts with pyruvate and His271α is involved in the reductive acetylation of the E2 lipoamide moiety (Pei *et al.*, 2008, Fig. 1-8B)

1.4.3.2 Dihydrolipoamide acetyltransferase (E2)

PDC E2 has a multidomain structure consisting of 1-3 lipoyl-domains (LD) at the N-terminus, followed by an E1/E3 sub-unit binding domain (SBD), and the catalytic domain (CD) at the C- terminus (Fig. 1-9A). The domains are linked together by 20-40 amino acid flexible linkers. E2 forms the so-called 'swinging arm' of PDC, which allows substrate channelling between E1, E2 and E3 (Perham and Reche, 1998). Eukaryotic E2 cores are formed with an additional protein, the E3 binding protein (E3BP), which has a multi-domain structure similar to that of E2, however, it is not catalytically active as it lacks a conserved histidine residue required for activity (Fig. 1-9B).

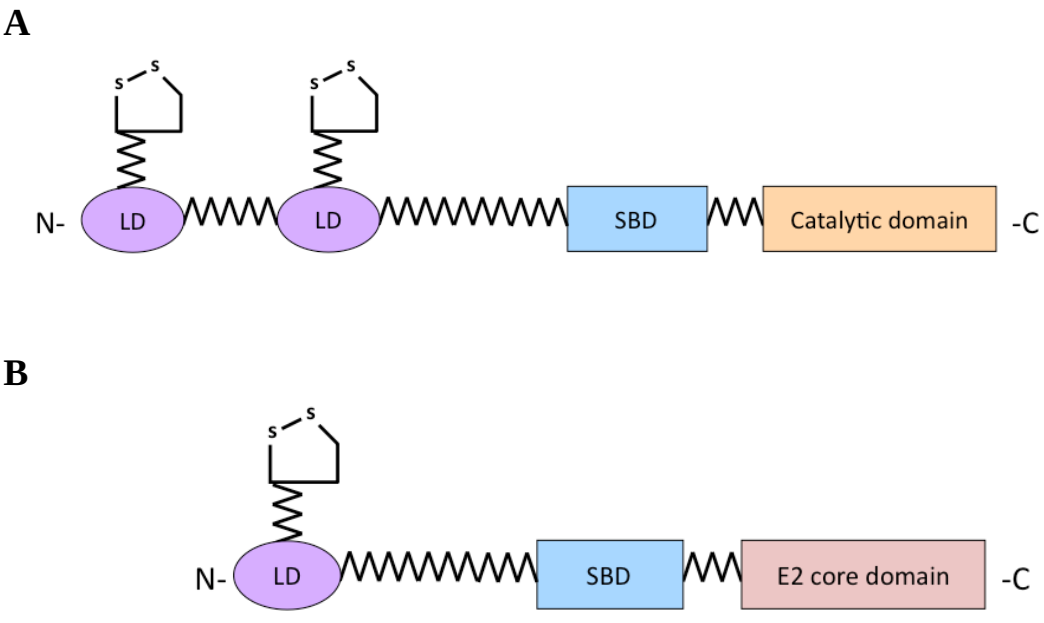


Figure 1-9. E2 multidomain structure
(A) E2 and (B) the E3 binding protein found only in eukaryotes. LD, lipoyl domain; SBD, sub-unit binding domain. Adapted from de Kok *et al*. (1998).

The lipoyl-domains contain a lysine residue, to which the co-factor lipoic acid binds via an amide linkage (de Kok *et al.*, 1998). The catalytic domain forms the core structure of the protein complex. PDC E2 can form into either a dodecahedral 60-mer as in humans, Gram-positive bacteria and plants or an octahedral 24-mer as in Gram-negative bacteria (Mattevi *et al.*, 1992a; Izard *et al.*, 1999; Mooney *et al.*, 1999; Yu *et al.*, 2008b, Fig. 1-10). PDC E2 monomers form a tightly bound trimer involving 48 hydrogen-bonds (16 per monomer), several van der Waals interactions and loops that extend over the surface of the neighbouring monomer (Mattevi *et al.*, 1992a). Conversely, the interactions involved in the large core formation are much weaker. A C-terminal 3_{10}-helix forms a "ball" that inserts into a socket formed by residues in a monomer of the neighbouring trimer resulting in four hydrogen bonds (2 bonds per ball and socket) (Mattevi *et al.*, 1992a).

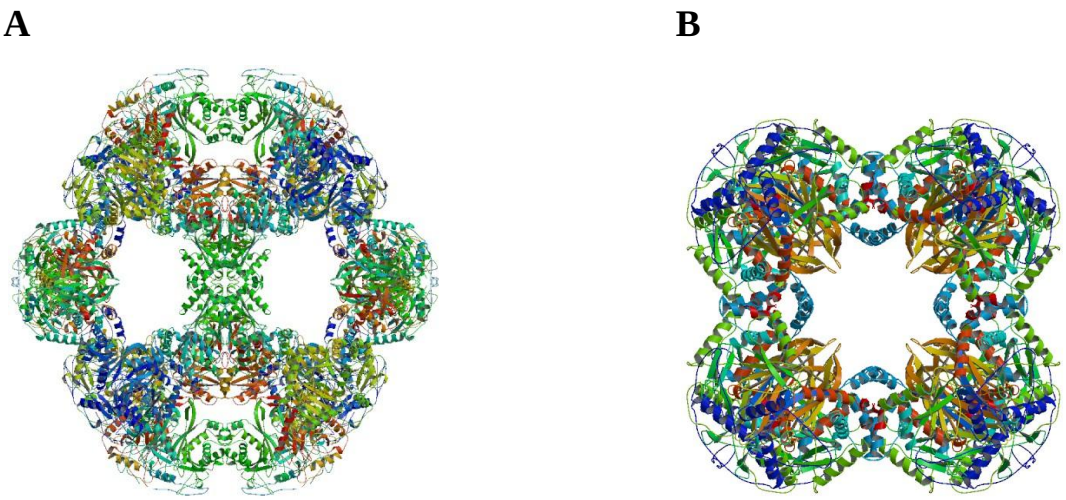


Figure 1-10. The core structure of PDC E2
(A) Icosahedral 60-mer PDC E2 core from *G. stearothermophilus* (formerly called *Bacillus stearothermophilus*, PDB entry 1B5S, Izard *et al*., 1999). (B) Octahedral 24-mer PDC E2 core form *Azotobacter vinelandii* (PDB entry 1EAA, Mattevi *et al.*, 1993a).

E2 catalyses the transfer of the acetyl group from S-acetyldihydrolipoamide to coenzyme A (CoA) producing acetyl-CoA (Fig. 1-11A). The sulphide carrying the acetyl group is reduced in the process. The active site is formed in the junction of two monomers. As an example, the active site residues in *A. vinelandii* are His610', Asn614' on one monomer (denoted from hereafter with the prime symbol) and Ser558 and Asp508 on the other monomer (Mattevi *et al.*, 1993a, Fig. 1-11B). In the inactive catalytic site Asp508 forms a hydrogen bond with Asn614' side chain preventing CoA from entering (Mattevi *et al.*,

1993a). Upon reduced CoA binding, Asn614' rotates and forms a hydrogen bond with His610' to allow the panthenine arm to enter the active site (Fig. 1-11C). This is followed by the E2 catalytic reaction where His610' extracts a hydrogen from the sulphydryl group of CoA (Mattevi *et al.*, 1993a, Hendle *et al.*, 1995). The sulphur then attacks the carbonyl carbon of the S-acetyldihydrolipoamide acetyl group forming a tetrahedral intermediate, which is stabilised by a hydrogen bond to Ser558 (Mattevi *et al.*, 1993a, Hendle *et al.*, 1995). The previously extracted hydrogen on His610' reduces the sulphur atom bound to the acetyl group, thus releasing acetyl-CoA and reduced dihydrolipoamide (Mattevi *et al.*, 1993a, Hendle *et al.*, 1995). Asn614' then rotates back to form a hydrogen bond with Asp508 to close the active site channel (Mattevi *et al.*, 1993a).

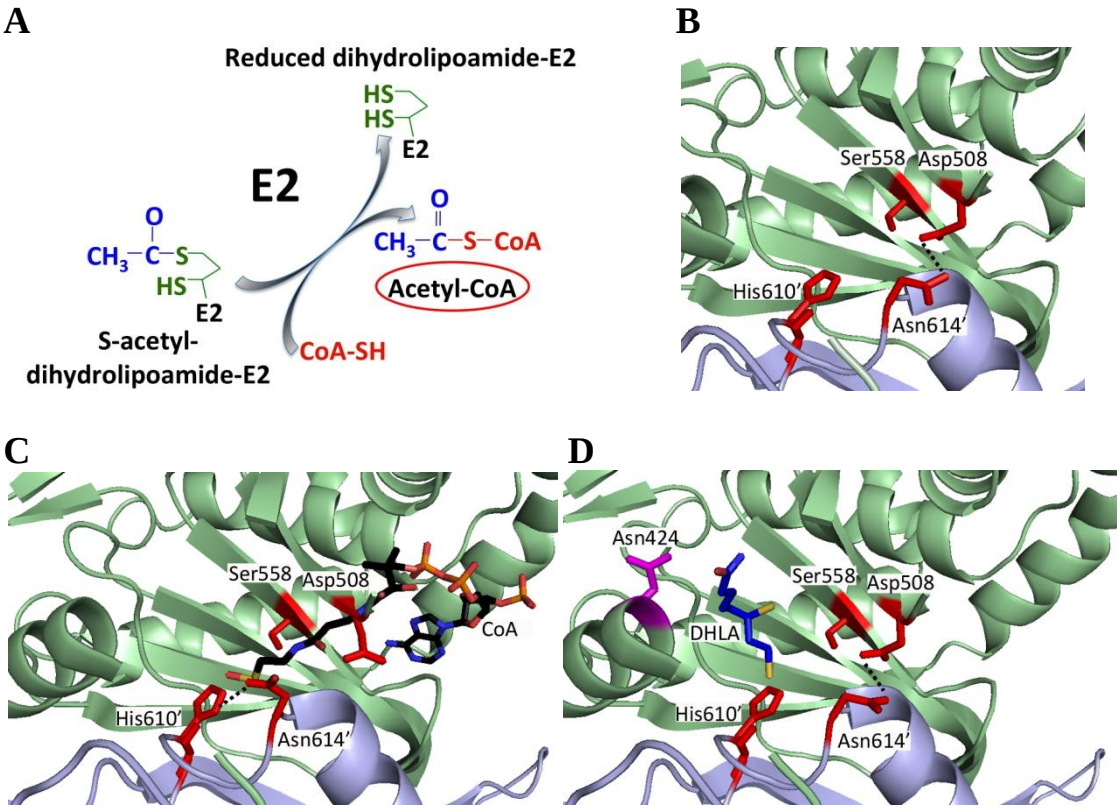


Figure 1-11. E2 reaction and active site
(A) E2 transfer of the acetyl group from S-acetyldihydrolipoamide to reduced CoA results in the formation of acetyl-CoA and the reduction of the lipoamide moiety of E2. (B) The active site of *A. vinelandii* E2 showing the "closed" conformation where Asn614' and Asp508 are involved in a hydrogen bond (dashed line) (PDB entry 1EAA, Mattevi *et al.*, 1993a). (C) The "open" conformation where Asn614' and His610' are involved in a hydrogen bond upon CoA binding (PDB entry 1EAD, Mattevi *et al.*, 1993a). (D) Co-crystallisation of E2 with dihydrolipoicacid (DHLA) showed the involvement of Asn424 in binding the co-factor (PDB entry 1EAE, Mattevi *et al.*, 1993a). Figures B-D adapted from Mattevi *et al.* (1993a).

A 30 Å channel at the interface of each of the E2 monomers forms the active site. The substrate CoA enters the active site from inside the core, while S-acetyldihydrolipoamide, is swung by the extended arm to the outer opening of the channel (Mattevi *et al.*, 1992a). The adenosine 3'phosphate moiety of CoA anchors the molecule to the protein allowing the movement of the panthenine arm (Fig. 1-11C), while dihydrolipoicacid (DHLA) binding involves a hydrogen bond with Asn424 (Fig. 1-11D) (Mattevi *et al.*, 1993a).

1.4.3.3 Dihydrolipoamide dehydrogenase (E3)

E3 is a flavoprotein disulphide oxidoreductase, a group of proteins characterised by the requirement for the co-factor flavin adenine dinucleotide (FAD) (Perham *et al.*, 1996). E3 also requires the co-factor nicotinamide adenine dinucleotide (NAD^+). For the formation of the active site E3 requires a homodimeric structure, which is highly conserved in bacteria, yeast and humans (Mattevi *et al.*, 1991; Mattevi *et al.*, 1993b; Toyoda *et al.*, 1998; Brautigam *et al.*, 2005, Fig. 1-12).

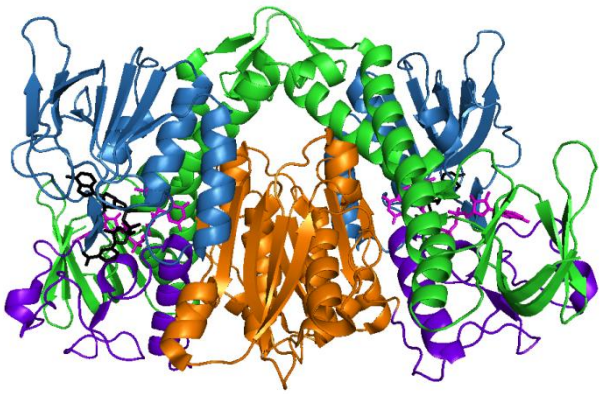


Figure 1-12. Structure of human E3
The FAD domain is in green, the NAD domain in blue, the central domain in purple and the interface domain in orange. FAD (magenta) and NAD (black) are shown as sticks. (PDB entry1ZMD, Brautigam *et al*., 2005).

A monomer of E3 has four domains. The residue numbers stated here are as reported for the human E3 by Brautigam *et al*. (2005). The largest domain is the FAD-binding domain formed by residues 1-149. The other three domains are the NAD^+-binding domain (residues 150-282), the central domain (residues 283-350) and the interface domain (residues 351-474). In addition to its role in the catalytic activity of E3, FAD binding is

required for the correct folding of the protein (Lindsay *et al.*, 2000). Thus, it is unsurprising that there are several residues forming a total of 22 hydrogen bonds to bind the co-factor to the enzyme (Mattevi *et al.*, 1991).

The E2 reaction leaves the sulphurs of the lipoamide cofactor reduced and these are re-oxidised by E3 to regenerate the co-factor for subsequent cycles of catalysis and to form NADH. E3 catalyses the oxidisation of the dihydrolipoamide moiety via two subsequent reactions, thus like E1, E3 exhibitis “ping-pong” kinetics (Fig. 1-13A).

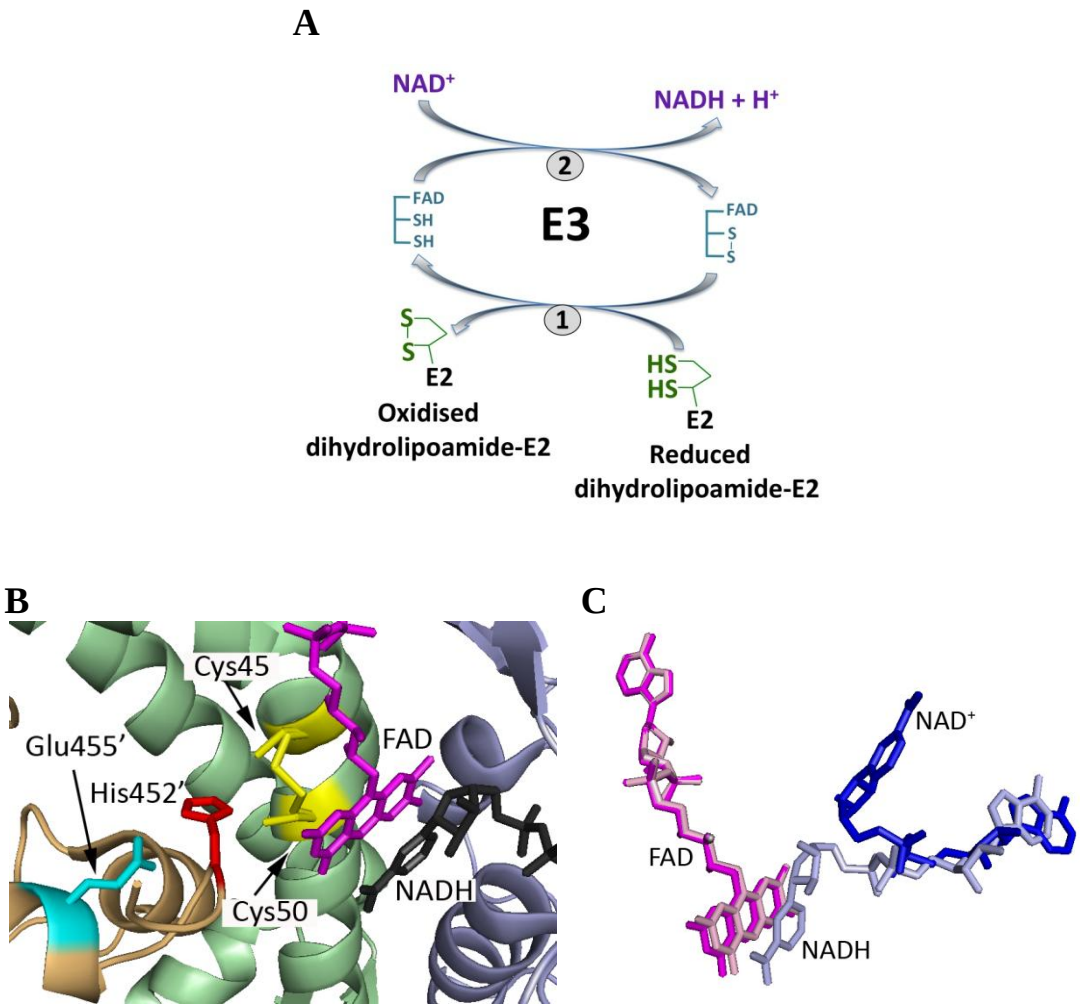


Figure 1-13. E3 reaction and active site
(A) E3 ping-pong kinetics. The first reaction (1) transfers a hydride ion from reduced dihydrolipoamide-E2 to FAD. The second reaction (2) transfers the hydride to NAD^+ forming NADH + H^+. (B) The human E3 active site residues and co-factors, see text for description. (C) Conformational change of NAD^+ (blue) NADH (light blue) upon active site formation. The conformation of the isoalloxazine ring while NAD^+ or NADH is bound is shown in magenta and light magenta, respectively. Figures adapted from Brautigam *et al*. (2005).

A hydride ion from the reduced lipoamide is transferred via a histidine residue onto a cysteine of the E3 active site disulphide. The cysteine forms a charge transfer complex with the isoalloxazine of FAD. NAD^+ then interacts with the isoalloxazine, facilitating the final transfer of the hydride to form NADH + H^+ (Mattevi *et al.*, 1991; Brautigam *et al.*, 2005).

The catalytic site involves a disulphide (formed by Cys45 and Cys50), His452' from the second monomer, which is held in the correct conformation by Glu455', and the isoalloxazine of FAD (Fig. 1-13B). NAD^+ is also present, however, it is not permanently bound to the active site. NAD^+ binding involves seven hydrogen bonds and there is a conformational change in the position of the nicotinamide ring during catalysis (Brautigam *et al.*, 2005, Fig. 1-13C). Following the formation of the charge-transfer complex by one of the active site cysteines and the FAD isoalloxazine on the *si*-side, the nicotinamide ring is aligned with isoalloxazine on the *re*-side to allow the transfer of the hyride (Brautigam *et al.*, 2005). The resulting NADH is bound to E3 by 12 hydrogen bonds (Brautigam *et al.*, 2005). Dimer formation involves nine hydrogen bonds, six of which are between residues in the interface domain and three in the FAD domain (Mattevi *et al.*, 1991). Residues involved in binding the SBD of E2 or the E3BP are also located in the interface domain.

1.4.3.4 The *P. falciparum* pyruvate dehydrogenase complex

P. falciparum is unusual in that the sole PDC is located in the apicoplast and not in the mitochondrion, as in humans and other eukaryotes (Foth *et al.*, 2005) or as in plants where a PDC is found in both the mitochondrion and plastid (Mooney *et al.*, 2002).

The *P. falciparum* PDC E1 enzyme is an $\alpha_2\beta_2$ heterotetramer. The E1α subunit (PlasmoDB: PF11_0256, www.plasmodb.org) is encoded on chromosome 11, the gene is 1824 bp and translates into a 608 amino acid protein with a predicted size of 70.9 kDa (Foth *et al.*, 2005; McMillan *et al.*, 2005). The gene for the E1β subunit (PlasmoDB: PF14_0441), on the other hand, is located on chromosome 14, is 1248 bp and encodes a protein with 416 amino acids with an estimated molecular mass of 47.0 kDa (Foth *et al.*, 2005; McMillan *et al.*, 2005). Amino acid sequence identity for E1α and E1β compared to the human equivalents are 32% and 35% respectively (McMillan, 2006). Both subunits

show the greatest amino acid identity (around 40-50%) to homologous proteins from cyanobacteria and red algae, which is consistent with the secondary endosymbiotic origin of the apicoplast (Foth *et al.*, 2005; McMillan, 2006). Although the protein sequence identity to homologues from other organisms is low, the PDC E1 subunits do contain conserved residues required for activity (Foth *et al.*, 2005).

The *P. falciparum* PDC E2 gene (PlasmoDB: PF10_0407) is 3192 bp, and it contains 10 predicted introns and complementary DNA analysis shows the full enzyme coding sequence is 1923 bp and the predicted protein consists of 641 amino acids, with an estimated molecular mass of 73.9 kDa (McMillan, 2006). The PDC E2 protein contains two lipoyl-domains, a SBD and a catalytic domain with conserved residues required for the acetylation reaction.

McMillan *et al*. (2005) showed that there are two E3 enzymes encoded by the nuclear genome of *P. falciparum*, and that one of these is targeted to the mitochondrion, the other to the apicoplast. The apicoplast PDC E3 gene (*aE3*) (PlasmoDB_PF08_0066) is located on chromosome 8, is 2001 bp and encodes a protein of 667 amino acids with a theoretical molecular mass of 75.6 kDa (McMillan, 2006). This differs from the mitochondrial E3 (mE3), which has a predicted molecular mass of 57.2 kDa, and is encoded by a 1539 bp gene (PlasmoDB: PFL1550w) on chromosome 12. The aE3 enzyme shows greatest sequence identity with cyanobacteria and plants, whereas the mE3 is similar to protists, *Dictyostelium*, plant and fungi homologues. This is unsurprising because of the different origins of the apicoplast and mitochondrion and in fact the aE3 and mE3 have an amino acid sequence identity of only 20%. Both proteins, however, do contain the conserved FAD, NADH, central and interface domains and the active site redox active cysteines (McMillan, 2006).

A recent study on the rodent parasite, *P. yoelii*, showed that the PDC is essential in the liver stage of the parasite. Deletion of *E1α* and *aE3* did not affect blood stage or mosquito stage parasite growth or development (Pei *et al.*, 2010). However, *E1α* and *aE3* were essential for the progression of the parasites from the liver stage to the infective blood stage in mice. Further investigation of *E1α* deletion mutants showed that specifically the late liver stages were affected. The mutants had growth and nuclear division defects and did not develop into mature schizonts and exo-erythrocytic merozoites (Pei *et al.*, 2010).

Also, the mutant parasites did not express merozoite surface protein 1 (MSP1), which is required for invasion of erythrocytes (Pei *et al.*, 2010). Comparable results were obtained with parasites lacking genes encoding proteins involved in FAS II, suggesting the major role of the PDC is to provide acetyl-CoA for FAS II (Yu *et al.*, 2008a; Vaughan *et al.*, 2009). In *P. berghei* and *P. falciparum, FabI* deletion did not affect blood stage parasites but the gene was important for development of *P. berghei* liver stage merozoites and *FabI* deletion resulted in reduced infectivity of mice (Yu *et al.*, 2008a). This effect was more pronounced in *P. yoelii*, where *FabB/F* and *FabZ* null mutants were shown to be unable to form exo-erythrocytic merozoites or to lead to a blood stage infection of mice (Vaughan *et al.*, 2009). Similar to the *P. berghei* study, FabB/F and FabZ were not required for erythrocytic stage growth or development in *P. yoelii* (Vaughan *et al.*, 2009). Deletion of the protein *N*-octanoyltransferase (LipB) also showed results similar to PDC and FASII mutant parasites. LipB is involved in producing the active form of E2 as it transfers the octanoyl-group from octanoyl-ACP, produced by the FASII pathway, to the lipoyl-domains of E2. This is followed by the addition of sulphurs to positions 6 and 8 of the octanoyl moiety by lipoic acid synthase (LipA). LipB was shown to not be essential for the blood stage growth and development of *P. falciparum* (Günther *et al.*, 2007). Deletion of the protein from *P. berghei* also showed it was not required for blood or mosquito stage growth. The *P. falciparum* LipB deletion mutants did, however, have a shorter cell cycle (Günther *et al.*, 2007), a phenotype that was not seen in the *P. berghei* mutants (Falkard *et al.*, 2013). Again, LipB was essential for successful late-stage liver development and the *P. berghei* mutant parasites had a reduced ability to infect mice erythrocytes (Falkard *et al.*, 2013). The reduced infectivity was likely to be due to residual lipoylation of PDC E2 (Falkard *et al.*, 2013), which could possibly be carried out by lipoic acid protein ligase-like protein (LplA2), a protein shown to be targeted to both the mitochondrion and apicoplast (Günther *et al.*, 2007). Together, these studies show that a functional PDC and FAS II are very important for the progression of the parasite from the asymptomatic liver stage to the symptomatic blood stage. Thus, the PDC could be a potential target for the development of prophylactic drugs or vaccines.

Even though the aforementioned studies suggest that the PDC is not required in the erythrocytic stages of *Plasmodium*, reverse transcriptase-PCR and western blotting have shown that the genes encoding the three subunits of PDC are expressed in *P. falciparum* during this stage (Foth *et al.*, 2005; McMillan *et al.*, 2005). RNA-seq data (available at www.plasmodb.org) show that the proteins are expressed at very low levels during the

intraerythrocytic cycle, however, higher levels are seen in late gametocytes and ookinetes (Bartfai *et al.* 2010; Lopez-Barragan *et al.*, 2011). In addition it was shown that PDC E2 is lipoylated and thus likely to be catalytically active during blood stage development (Günther *et al.*, 2007). Because the PDC is expressed, questions arise as to whether the enzyme has another role in the parasite. For example, E2 and E3 may play a role in redox regulation via the covalently attached lipoyl-group on E2 and the NADH/NAD^+ cycling carried out by E3. Free lipoic acid is a potent antioxidant with therapeutic potential. The reduced/oxidised lipoic acid redox couple has metal chelating properties, can scavenge reactive oxygen species (ROS), regenerate other antioxidants such as vitamin C, vitamin E and glutathione and also plays a role in repair of oxidatively damaged proteins (Biewenga *et al.*, 1997). Therapeutical lipoic acid has been suggested for aging, cancer and diabetes (Goraca *et al.*, 2011). Protein bound lipoic acids may also have a role outside their function as acetyl/succinyl transferases in the KADC complexes. In *Mycobacterium tuberculosis*, peroxiredoxin alkyl hydroperoxide reductase (AhpC) is linked to the E2 enzyme of the KGDC complex and E3 via a thioredoxin-like protein (AhpD) to form a NADH-dependent peroxidase and peroxynitrite reductase (Bryk *et al.*, 2002). This complex is important in *M. tuberculosis* defence against oxidative and nitrosative stress exerted by the host cells following an immune response to the bacterium. Also, the presence of several lipoylated domains on E2, all of which are not required for sufficient catalytic activity (Dave *et al.*, 1995; Guest *et al.*, 1997), suggests that the lipoic acid moieties may interact with other molecules in the surrounding environment (Bunik, 2003). E3 has several redox active roles. It has been implicated in NADH-dependent scavenging of nitric oxide (Igamberdiev *et al.*, 2004) and reduction of ubiquinone to the antioxidant ubiquinol (Olsson *et al.*, 1999; Xia *et al.*, 2001). E3 has also been shown to protect the mammalian E2-E3 binding protein core from lipid peroxidation products (Korotchkina *et al.*, 2001). Thus, one of the aims of this part of the project was to investigate, whether the *P. falciparum* PDC could play a role in oxidative stress encountered by the parasite.

Another important factor is that acetyl-CoA is not only required for metabolic pathways but it is also a major source of acetyl groups for protein acetylation (Xing and Poirier, 2012). Acetylation has traditionally been associated with histones but recently it has been discovered to also be involved in a vast range of functions outside the nucleus and has been realised to be an important post-translational modification comparable to phosphorylation (Norris *et al.*, 2009). Acetylation occurs at least in humans, plants and bacteria (Choudhary *et al.*, 2009; Jones and O'Connor, 2011; Xing and Poirier, 2012).

Acetyl-CoA plays a role in three types of acetylation including O-acetylation, the addition of an acetyl group to the hydroxyl group of an internal serine or threonine, and Nα-acetylation, the addition of an acetyl group to the N-terminal amino acid of a protein (Xing and Poirier, 2012). The most common form is Nε-acetylation, in which an acetyl group is added to the side-chain amino group of an internal lysine, thus also called lysine acetylation (Xing and Poirier, 2012). This acetylation is reversible and mediated by lysine acetyl transferases (KATs) and lysine deacetylases (KDACs) (Xing and Poirier, 2012). Thousands of proteins were found to be acetylated in human cells and acetylation-mediated networks were involved in RNA splicing, DNA damage repair, the cell cycle, nuclear transport and actin cytoskeleton remodelling (Choudhary *et al.*, 2009). Acetylated proteins also play a role in other cellular processes such as proteolysis, vesicular trafficking, endocytosis and autophagy, stress response and even some kinases and phosphatases are acetylated (Choudhary *et al.*, 2009; Jones and O'Connor, 2011; Xing and Poirier, 2012). *P. falciparum* has been shown to encode at least four KATs as well as several KDACs and histone deacetylases (HDACs) (Miao *et al.*, 2010; Miao *et al.*, 2013). Acetylation in *P. falciparum* has mainly been studied in context of histone acetylation (Miao *et al.*, 2006; Goyal *et al.*, 2012) and potential drug candidates against parasite HDACs have been researched (Wheatley *et al.*, 2010; Marfurt *et al.*, 2011). There are also reports of Nα-acetylation of *P. falciparum* exported proteins and actin (Schmitz *et al.*, 2005; Chang *et al.*, 2008). A recent study has been carried out on *P. falciparum* lysine acetylation, which identified 230 proteins that were acetylated (Miao *et al.*, 2013). Most of these proteins were cytosolic (n = 167) but ten were shown to be plastidial (Miao *et al.*, 2013). The *P. falciparum* acetylated proteins are mainly involved in translation, transcription, metabolism and chaperones and not as much in replication, invasion and structural functions (Miao *et al.*, 2013). Clearly as acetylation is very important in the parasite and because acetyl-CoA is a major contributor of acetyl-groups, the role of the PDC in producing acetyl-CoA for acetylation would be an appealing perspective to study in *P. falciparum*.

In conclusion, because the mutation of the PDC has been shown to inhibit egress of parasites from hepatocytes, thus preventing the initiation of the symptomatic intraerythrocytic cycle, the PDC could be a potential new target for prophylactic drugs and vaccines. This warrants further investigation into the structure and function of the *P. falciparum* PDC to determine whether there are any differences between the parasite and

human PDC, which could be exploited for drug development. In addition, studying the role of the PDC in the intraerythrocytic stages of *P. falciparum* will be very interesting to further understand how the parasite functions.

1.5 Aims of the project

The specific questions this project will address are:

a. What are the structures and stoichiometries of the *P. falciparum* PDC components?
b. What are the kinetic properties of the enzymes?
c. Are there any biochemical and structural differences in the *P. falciparum* PDC compared with the human complex that are exploitable for future drug development?
d. Is the PDC required for *P. falciparum* blood stage development?

Hypotheses

1. The *P. falciparum* PDC differs biochemically from the human PDC.
2. The *P. falciparum* PDC differs structurally from the human PDC.
3. The *P. falciparum* PDC is not essential for the parasite blood stage development.

Testing these hypotheses will demonstrate drug target suitability of the *P. falciparum* PDC.

1.6 Introduction to biophysical techniques

1.6.1 Analytical ultracentrifugation (AUC)

In 1924, Theodor Svedberg built his first successful analytical ultracentrifuge (van Holde and Hansen, 1998). At a time when proteins were thought to be merely aggregated colloids of small molecules, his first studies on haemoglobin were monumental in protein research (Rosenfeld, 1982). By performing sedimentation equilibrium (SE) experiments, Svedberg and Fahraeus (1926) showed that the molecular mass of haemoglobin was 68,000 Da, which was exactly four times the minimum molecular mass (16,700 Da) estimated for the protein based on its iron content. For the first time it was shown that proteins were in fact specific molecules with large molecular masses that formed organised macromolecules rather than the previously thought aggregated clusters of small particles. Svedberg then went on to develop an oil turbine ultracentrifuge, which was capable of producing rotor speeds fast enough to measure sedimenting boundaries, thus developing sedimentation velocity (SV) analysis (Svedberg and Nichols, 1927). The first SV experiment was also carried out on haemoglobin and corroborated the SE results, showing a homogenous sample with a molecular mass of 68 kDa (Svedberg and Nichols, 1927).

The development of the more powerful Spinco Model E analytical ultracentrifuge made AUC a popular method to study not only protein but also other molecules such as DNA (Bauer and Pickels, 1937). In the late 1970's, however, the use of AUC almost stopped completely as the Spinco Model E became outdated and cheaper, easier methods such as gel filtration and gel electrophoresis were developed and were used to approximate protein molecular mass and stoichiometry (van Holde and Hansen, 1998). The next major development was not until 1992, when Beckman-Coulter (Palo Alto, USA) introduced the much-improved XL-A analytical ultracentrifuge with scanning absorption optics and digital data acquisition, followed by the XL-I in 1996, which has both absorption and Rayleigh interference optics that can be used at the same time (Schuster and Toedt, 1996; van Holde and Hansen, 1998). The development of more powerful computers and software also facilitated faster analysis of data (van Holde and Hansen, 1998). Following these developments AUC has yet again become a powerful tool in biomedical research.

AUC has several applications in biological research. First of all, it is an unrivalled method of directly determining the precise molecular mass of any molecule in their near native state in solution without relying on calibration standards (Ralston, 1993). Secondly, sample purity can be examined using AUC. SV analysis shows how many species there are and how much of each there is in a sample and SE analysis can be then used to determine the molecular mass of each species (Ralston, 1993). In addition, sedimentation analyses can be used to study associating systems. The change in molecular mass following complex formation or dissociation can be measured and the strength of the interaction can be calculated (Ralston, 1993). Other applications include determining the thermodynamic non-ideality of a solution, measuring conformational changes of proteins and ligand binding studies (Ralston, 1993).

1.6.1.1 Theory for sedimentation of particles

There are three forces acting on sedimenting particles (Fig. 1-14). The following is the description of the mechanical model of sedimentation as described in Ralston, (1993).

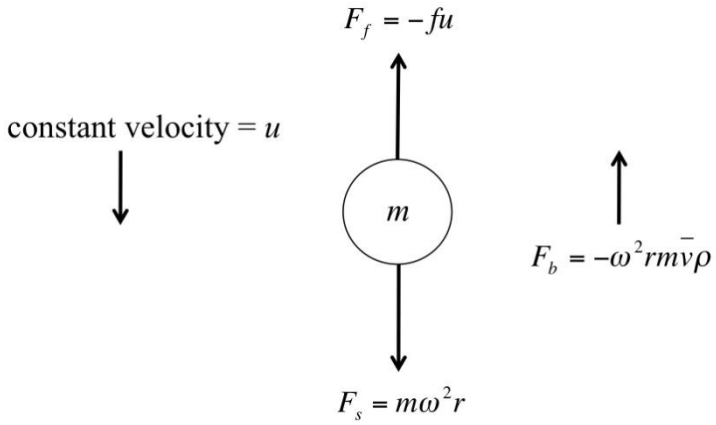


Figure 1-14. Forces acting on sedimenting particles
When centrifugal force (constant velocity) is applied to a molecule the sedimentation force (F_s) is opposed by the frictional force (F_f) and buoyancy (F_b). Figure adapted from Ralston, (1993).

The sedimenting, or gravitational, force, F_s, depends on the mass of the particle m, and its linear acceleration, which in a rotating system is defined as the squared angular velocity (in radians per second) ω^2, multiplied by the distance from the centre of rotation (r)

$$F_s = m\omega^2 r = \frac{M}{N}\omega^2 r \qquad (1.4)$$

where m is derived from the molar mass M of the molecule (g/mol) and Avogadro's number N (6.023 x 10^{23} /mol). A second force, buoyancy, F_b, acts in the opposite direction of sedimentation and which is given by Archimedes' principle 'the buoyancy of a body in liquid equals the weight of the liquid in a volume equal to the volume of the body' (Serdyuk *et al.*, 2007a). Thus, F_b is defined as

$$F_b = -m_0\omega^2 r \qquad (1.5)$$

where $\omega^2 r$ is the linear acceleration of the particle (the minus in $-m_0$ denotes the opposite direction of F_b compared with F_s) and m_0 is

$$m_0 = m\bar{v}\rho = \frac{M}{N}\bar{v}\rho \qquad (1.6)$$

where $\bar{v}$ is the partial specific volume (ml/g) of the particle and ρ the density of the liquid (g/ml) in which the particle is immersed. The partial specific volume is defined as the change in volume when 1 gram of a molecule is solubilised.

When the density of the particle is greater than the density of the liquid, the particle will start sedimenting towards the bottom of the cell. The velocity, u, will initially for a fraction of time increase the further away it moves along the radius from the centre of rotation. The viscosity of the liquid will cause friction and affect the velocity. Frictional force, F_f, is the third force affecting sedimentation and is defined as

$$F_f = -fu \qquad (1.7)$$

where f is the frictional coefficient and the minus indicates the opposite direction of F_f compared with F_s. The frictional coefficient depends on the shape and size of the sedimenting molecule; f for smooth and compact spheres is smaller than for bulky and elongated particles.

When the three forces come into equilibrium (usually in less than 10^{-6} s) constant velocity is achieved

$$F_s + F_b + F_f = 0 \qquad (1.8)$$

$$\frac{M}{N}\omega^2 r - \frac{M}{N}\bar{v}\rho\omega^2 r - fu = 0 \qquad (1.9)$$

which can be rearranged to

$$\frac{M}{N}(1-\bar{v}\rho)\omega^2 r - fu = 0 \qquad (1.10)$$

This is further rearranged so that the experimental parameters are on the left and the molecular parameters on the right (Serdyuk *et al*., 2007a)

$$s \equiv \frac{u}{\omega^2 r} = \frac{M(1-\bar{v}\rho)}{Nf} \qquad (1.11)$$

As sedimenting particles move toward the bottom of the cell, diffusion of the particles occurs, which broadens the boundary. The rate of the spreading boundary is determined as the diffusion coefficient, *D*, and depends on the size of the molecules

$$D = \frac{RT}{Nf} \qquad (1.12)$$

where *R* is the gas constant, *T* the absolute temperature and *N* Avogadro's number. Adding the diffusion constant to the equation determining the sedimentation coefficient, gives Svedberg's equation (Lebowitz *et al.*, 2002):

$$s = \frac{u}{\omega^2 r} = \frac{M(1-\bar{v}\rho)}{Nf} = \frac{MD(1-\bar{v}\rho)}{RT} \qquad (1.13)$$

From this we can determine the *sedimentation coefficient*, *s*, which is the ratio of the velocity of the particle to its centrifugal acceleration ($u/\omega^2 r$). The dimensions of *s* are seconds. The sedimentation coefficient for biological molecules are around 1-100 x 10^{-13} s,

thus the basic unit is 10^{-13} s and is called a Svedberg unit (S) in acknowledgment of Theodor Svedberg's work in AUC. Molecules of different shapes and sizes will move at different velocities when centrifuged and will thus have different sedimentation coefficients.

To be able to compare sedimentation coefficients from different experiments conducted at different temperatures and in differing buffers, the experimental values ($s_{T,B}$) are converted to $s_{20,w}$, i.e. standard values at 20°C in water (w):

$$s_{20,w} = s_{T,B}\left(\frac{\eta_{T,B}}{\eta_{20,w}}\right)\frac{(1-\bar{v}\rho)_{20,w}}{(1-\bar{v}p)_{T,B}} \qquad (1.14)$$

where T is experimental temperature and B is the experimental buffer conditions, η is viscosity of water or buffer (Poise), $\bar{v}$ is the partial specific volume (ml/g) of the studied protein and ρ the density of water or buffer (g/ml) (Lebowitz *et al.*, 2002).

1.6.1.2 AUC experiment: instrumentation and optics

For an AUC experiment, a protein sample and the reference buffer are loaded into a double sector centrepiece flanked by optical-grade quartz or sapphire windows. This is placed into a metal housing and the cell is placed into an anodised titanium rotor. The samples are centrifuged at 3,000 to 60,000 rpm in a vacuum chamber, designed to decrease frictional heating and aerodynamic turbulence. The Beckman Coulter (Palo Alto, USA) Optima XL-I analytical ultracentrifuge uses both absorbance and Rayleigh interference optics. Data in this thesis were mainly collected using the absorbance optics. The absorbance optics operate at wavelengths ranging from 190 – 800 nm (Fig. 1-15). For experiments on proteins, the absorbance can be set to 280 nm to detect the aromatic side chains of amino acids or 230 nm, the wavelength absorbed by the peptide bonds (Ralston, 1993). Rayleigh interference optics produce fringe patterns, which arise from light passing slower though the sample compared with the reference buffer. The difference in the refractive index of the sample and reference buffer, which is proportional to the concentration difference, causes a vertical shift in the fringe pattern in the area of the sedimenting protein boundary (Ralston, 1993). The absorbance optics are valuable for measuring low concentrations of

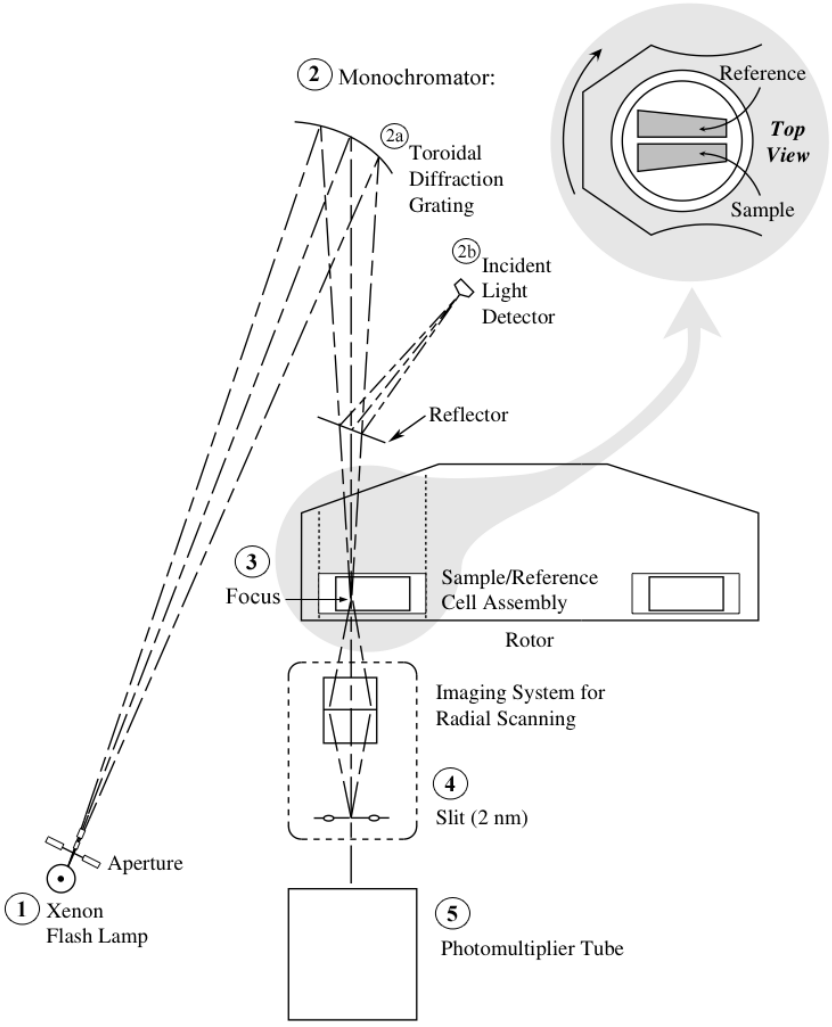


Figure 1-15. Absorbance optical system
A schematic of the absorbance optical system of the Beckman Optima XL-A analytical ultracentrifuge. The optics works the same way in the Beckman Optima XL-I analytical ultracentrifuge used in this study. The xenon lamp (1) flashes each time a measurement is taken. The light is directed via a monochromator (2), which contains a torroidal diffraction grating (2a) that selects the correct wavelength and collimates the light. The incident light detector (2b) in the monochromator measures and normalises the intensity of each flash of light before it reaches the sample. (3) The light is focused to 1/3 of the cell sector depth. The 2 nm slit (4) moves beneath the sample cell to allow measurements at different radial positions. The photomultiplier tube (5) measures the intensity of the light during a scan. Figure taken from Ralston (1993), with permission from Beckman Coulter.

proteins because to keep within the linear range of Beer's Law, absorbance of the sample should not exceed 1.0 (Schirf and Planken, 2008). Rayleigh interference is useful when measuring high concentrations of proteins when the absorbance would be too high, and also for substances that do not absorb light, for example polysaccharides (Schirf and Planken, 2008).

1.6.1.3 Sedimentation velocity (SV)

In a SV experiment samples are centrifuged at 3,000 – 49,000 rpm depending on the size of the protein. Large macromolecules need to be centrifuged at lower speeds to be able to monitor their movement; at high speed the protein would sediment too fast. Likewise small molecules need higher speeds to force sedimentation. It is the movement of the boundary region (Fig. 1-16), which is monitored and recorded for analysis.

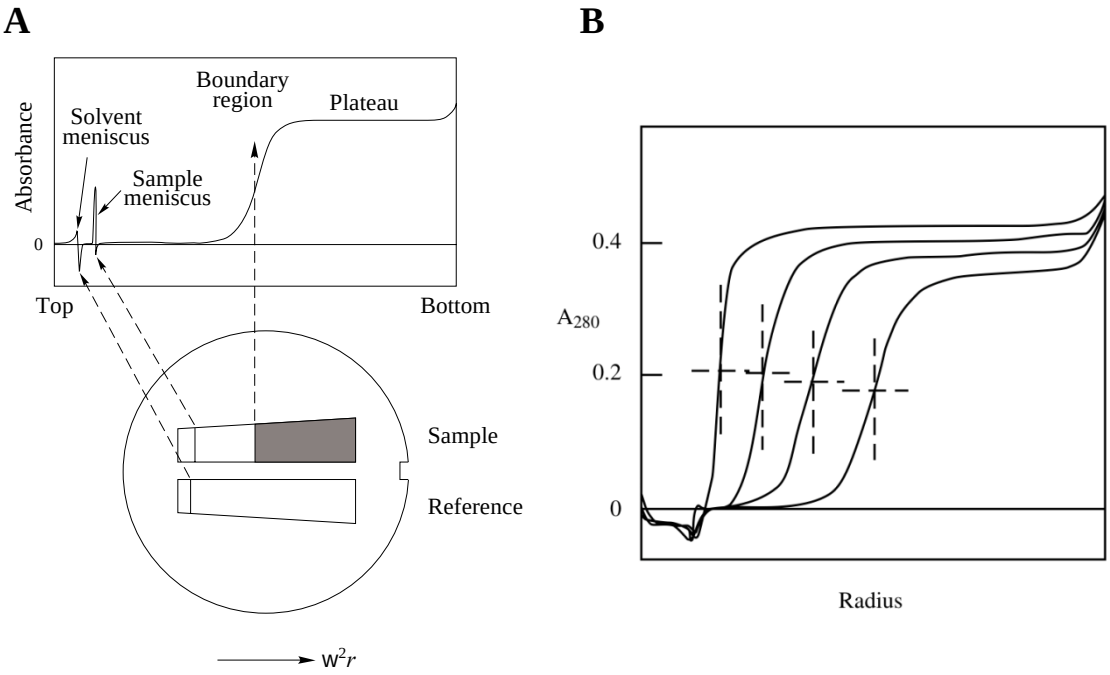


Figure 1-16. Sedimenting boundaries in a SV experiment
(A) Diagram showing the AUC sample cell and one recorded fringe profile. (B) The boundary region moves as the proteins sediment and, in the simplest analysis, this movement is used to obtain the sedimentation coefficient. Figures taken from Ralston (1993), with permission from Beckman Coulter.

The Lamm equation is used for direct modelling of the movement of sedimenting boundaries. The Lamm equation is a partial differential, which describes the sedimentation and diffusion of the boundary of a monodisperse solution in a sector shaped cell (Schuck *et al.*, 2002)

$$\frac{\partial \chi(r,t)}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left[rD\frac{\partial \chi(r,t)}{\partial r} - s\omega^2 r^2 \chi(r,t) \right] \qquad (1.15)$$

where χ is the concentration of the sedimenting particle, *r* the radial position in the sector shaped cell and *t* time. The first part of the equation in the brackets describes diffusion (*D*) and the second part sedimentation (*s*) of the particle. In this thesis, the programme SEDFIT (Schuck, 2000; Schuck *et al.*, 2002) was used to obtain the sedimentation coefficient of proteins. Amongst other models, this programme generates the differential concentration distribution c(*s*) using the solution of the Lamm equation to model the experimental absorbance or interference profiles with the following equation (Schuck *et al.*, 2002)

$$a(r,t) = \int c(s)\chi(s,D(s),r,t)ds + \varepsilon \qquad (1.16)$$

where $a(r,t)$ is the experimental sedimentation profile, c(*s*) is the concentration distribution of sedimentation coefficients between *s* and *s*+*ds*, $\chi(s,D(s),r,t)$ is the solution of the Lamm equation and ε is error. By integrating the area under the peaks the weight-average sedimentation coefficient can be determined for each species in a sample. This analysis does not require prior knowledge of how many species are present in the sample and is not constrained in how many species it can fit. Thus c(*s*) analysis was used to determine sedimentation coefficients in this thesis.

1.6.1.4 Sedimentation equilibrium (SE)

During an SE experiment molecules are centrifuged at lower speeds than in an SV experiment. The lower centrifugal speed allows diffusion of the protein to oppose the sedimentation force (Lebowitz *et al.*, 2002). Following centrifugation for one or two days (dependent on molecular mass) these two processes reach equilibrium and the protein forms a concentration gradient in which no net movement of molecules occurs (Fig. 1-17). The equilibrium distribution is dependent on the molecular mass of the protein, thus its mass can be determined from SE data (Ralston, 1993).

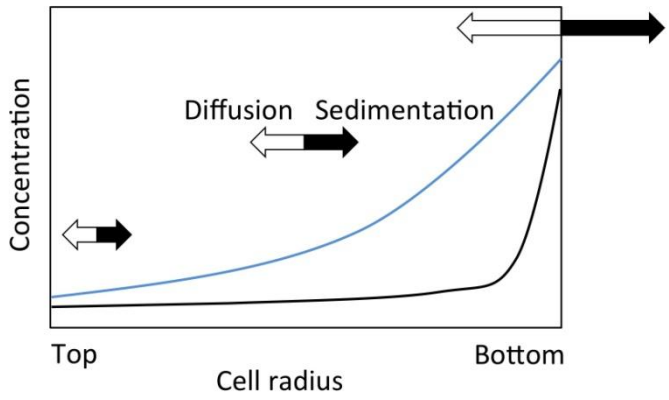


Figure 1-17. Sedimentation equilibrium experiment
The opposing forces, sedimentation and diffusion, act on proteins until a concentration gradient is achieved where no net movement is observed. The steepness of the gradient curve depends on the size of the protein; large proteins will pack towards the bottom of the cell resulting in steep curves (black) and vice versa for smaller proteins (blue). The length of the arrows denotes the increasing sedimentation force and diffusion, which occurs towards the base of the cell. Figure adapted from Ralston, 1993.

Sedimentation is proportional to $\omega^2 r$ and r increases towards the bottom of the cell, thus the sedimentation force increases towards the bottom of the cell (Ralston, 1993). Therefore, there must be a higher degree of diffusion at the bottom of the cell to counteract the sedimentation force. Diffusion is dependent on the concentration gradient, thus the protein concentration increases towards the bottom of the cell (Ralston, 1993). For an ideal non-associating single species the following equation can be used to determine the molecular mass (M) from the concentration gradient

$$M = \frac{2RT}{(1-\bar{v}\rho)\omega^2} \times \frac{d(\ln c)}{dr^2} \qquad (1.17)$$

where c is the concentration of the protein (Ralston, 1993). Thus, for a single ideally sedimenting species at equilibrium a slope proportional to the molecular mass can be obtained by plotting the log concentration against the radius squared (Ralston, 1993).

For a mixture of non-interacting ideal species the following equation can be used (Lebowitz *et al.*, 2002)

$$a(r) = \sum_n c_{n,0}\varepsilon_n d \exp\left[\frac{M_n(1-\bar{v}_n\rho)\omega^2}{2RT}(r^2 - r_0^2)\right] + \delta \qquad (1.18)$$

where $a(r)$ is the measured signal as a function of the radial position, $c_{n,0}$ is the molar concentration of species n at position r_0, M_n is the molecular mass, $\bar{v}_n$ is the partial specific volume, ε_n is the molar extinction coefficient, d is the optical path length of the sector (1.2 cm), and δ is the baseline offset, which compensates small imperfections due to cell assembly, data acquisition and differences between protein sample and buffer (Lebowitz *et al.*, 2002). SE data were analysed using the programme SEDPHAT (Schuck, 2003; Vistica *et al.*, 2004), which uses the above equations amongst others to fit the experimental data. Analyses in this thesis were carried out using the single species non-interacting discrete species model to obtain the apparent whole-cell weight average molecular mass ($M_{w,app}$). A two or three species non-interacting discrete species model analysis was used to determine the molecular mass of each of the species in samples.

While the equilibrium distribution is dependent on the molecular mass of an ideal single species protein, thermodynamic non-ideality due to elongated shape and/or the charge carried by proteins can complicate SE analysis (Ralston, 1993). This will affect the excluded volume of proteins, which for ideal globular proteins will be smaller than for non-ideal proteins, which in turn will affect how the molecules pack towards the base of the cell during an SE experiment. Decreasing $M_{w,app}$ values as the concentration of the protein increases is indicative of thermodynamic non-ideality (Ralston, 1993). The magnitude of non-ideality can be determined from the second virial coefficient, which can be determined using the equation

$$M_{app} \approx \frac{M}{(1+BMc)} \qquad (1.19)$$

where B is the second virial coefficient (Ralston, 1993). Higher B values indicate greater thermodynamic non-ideality.

1.6.2 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) was first developed by André Guinier in the late 1930's when he showed that X-ray scattering by particles at low angles contained shape and size information (Serdyuk *et al.*, 2007b). SAXS has been used to study biomolecules in solution since the 1960's with major advances in the 1970's with the development of brighter synchrotron radiation (Svergun, 2007). Today, third-generation synchrotrons allow scattering from molecules with very short exposure times (seconds or less) to limit radiation damage to samples (Svergun, 2007). The major challenge in SAXS remains the accurate analysis of the low resolution data to obtain high quality three-dimensional structural information (Svergun, 2007).

SAXS is a very useful tool to study proteins and other macromolecules such as DNA and RNA in their native state in solution. Recent advances in instrumentation and software for data analysis have enabled SAXS data to be used to obtain information about protein quaternary structure and flexibility, acquire low resolution (10 - 20 Å) *ab initio* models and calculate theoretical scattering curves for known crystal structures or homology models (Mertens and Svergun, 2010). The amount of protein required for an experiment depends on the synchrotron used. At second-generation beamlines, such as DORIS III at the European Molecular Biology Laboratory (EMBL)/Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany, 1-2 mg of protein is required. At third-generation synchrotrons, such as PETRA III (also at DESY), measurements can now be done on as little as 30 μg of protein. The sample does need to be of high quality, ideally homogenous and not aggregate in solution for best results (Mertens and Svergun, 2010). Virtually any size of protein can be studied ranging from small 1-10 kDa polypeptides to large MDa viral capsids and sample environment effects (pH, salt concentration, temperature, addition of ligands, co-factors and other proteins) can easily be measured (Mertens and Svergun, 2010). In cases where crystal structures cannot be obtained, SAXS provides important structural information. SAXS based solution structure envelopes can also be compared with crystal structures to see whether the native state of the protein is retained following crystallisation. The more advanced programmes available can also model missing structures, protein flexibility and oligomeric mixtures of proteins (Petoukhov *et al.*, 2012).

1.6.2.1 SAXS instrumentation

SAXS experiments were carried out at the X33 beamline at the EMBL/DESY, Hamburg, Germany. The beamline has since been decommissioned, however, X33 was a bending magnet beamline on the DORIS III storage ring, which operated at beam currents of 100 - 140 mA with an X-ray wavelength of 0.15 nm and an electron energy of 4.5 GeV. The beam was directed through several slits and an X-ray mirror, which focused the beam and eliminated stray scattering (Roessle *et al.*, 2007, Fig. 1-18). The beam was then passed through the sample and the data were collected on a 2D photon counting Pilatus 1M-W pixel X-ray detector positioned 2.6 m from the sample (Blanchet *et al.*, 2012). The raw data were then fed into an automated SAXS data processing pipline (Petoukhov *et al.*, 2007) for analysis of results as described in section 2.7.4.

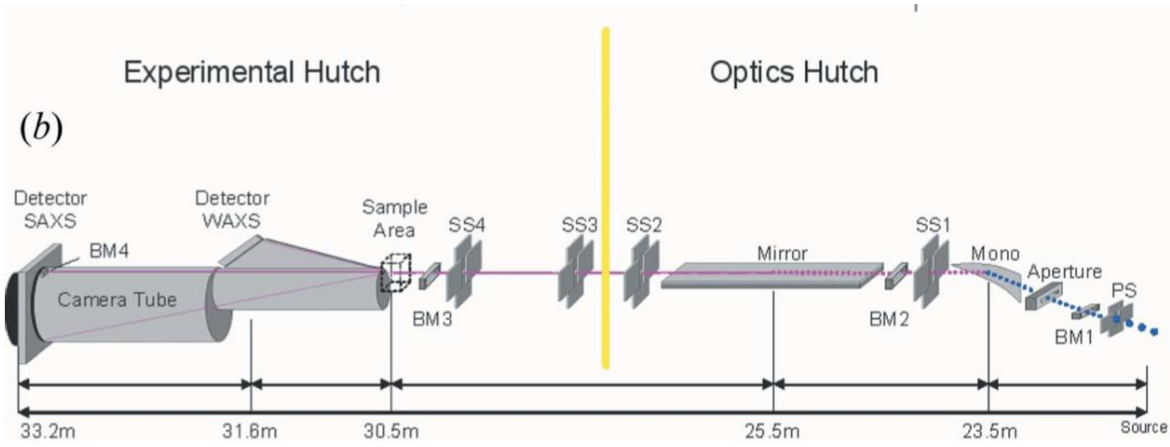


Figure 1-18. SAXS beamline
The beam passes though a primary slit (PS), the monochromator (Mono) and the first of several secondary slits (SS1) and is then focussed with the X-ray mirror. Before the beam is directed into the sample, parasitic scattering is eliminated with three more slits (SS2-4), two of which are in the experimental hutch. Scattering data can then be collected on a wide-angle X-ray scattering detector (WAXS) or a SAXS detector. The beam quality is monitored at several points (BM1-3). Figure taken from Roessle *et al.*, 2007 with permission from the International Union of Crystallography (IUCr).

1.6.2.2 Theory of SAXS

The scattering of X-rays by particles described here is based on the molecules being at infinite dilution, i.e. in conditions where no interparticle interactions take place. A schematic of SAXS is shown in Fig. 1-19.

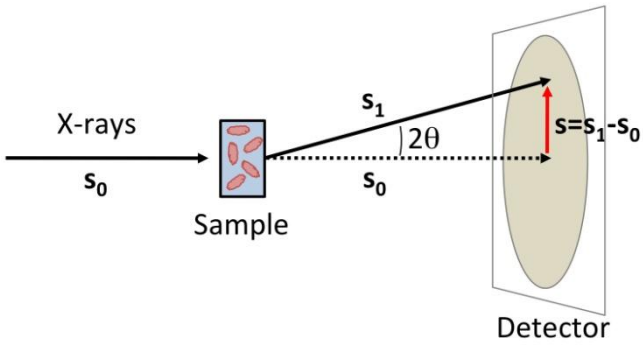


Figure 1.19. Scattering of X-rays in SAXS
Schematic representation of a SAXS experiment where the incident X-rays ($\boldsymbol{s_0}$) are scattered ($\boldsymbol{s_1}$) by macromolecules in solution at an angle of 2θ. This results in the scattering vector $\boldsymbol{s}=\boldsymbol{s_1}-\boldsymbol{s_0}$. Figure adapted from Mertens and Svergun, 2010.

At infinite dilution all the scattered particles are identical and will have the same isotropic scattering amplitude (Serdyuk *et al.*, 2007c; Mertens and Svergun, 2010). The scattering intensity (I(s)) is the sum of the scattering from all the particles (N) in the solution taking into account the different orientation of each particle (Serdyuk *et al.*, 2007c).

$$I(s) = N\left\langle \left|F(s)\right|^2 \right\rangle \qquad (1.20)$$

where $\boldsymbol{s}$ is the scattering vector given by $\boldsymbol{s}=\boldsymbol{s_1}-\boldsymbol{s_0}$ and its magnitude is $4\pi\sin(\theta)/\lambda$ (where 2θ is the angle between the incident and scattering beam, and λ is the wavelength, Fig. 1-19) and N is the number of particles in the solution. The latter part of the equation is explained as follows. A wave that has been scattered by a single particle is expressed as a function of the scattering vector (Serdyuk *et al.*, 2007c)

$$F(s) = \sum_j f_j \exp(is\cdot r_j) \qquad (1.21)$$

where f_j is the scattering amplitude and r_j the position of atom j. The scattering intensity is the wave amplitude squared (Serdyuk *et al.*, 2007c)

$$I(s) = |F(s)|^2 \qquad (1.22)$$

As each particle in the solution will scatter X-rays at different orientations, the $\langle ... \rangle$ brackets in equation 1.21 denote the averaged orientation of the particles (Serdyuk *et al.*, 2007c). The scattering intensity of proteins in solution also depends on the amount of protein in the solute, which is taken into account with the excess scattering length density, $\Delta p(r)$, the difference between the electron density of the protein and the solute (Mertens and Svergun, 2010). Thus the final recorded scattering intensity, $I(s)$, is

$$I(s) = \langle \mathbf{I(s)} \rangle_\Omega = \langle \mathbf{A(s)A*(s)} \rangle_\Omega \qquad (1.23)$$

where Ω denotes scattering intensity over all orientations and **A**(**s**) is the Fourier transformation of the excess scattering length density (Mertens and Svergun, 2010). Finally, the buffer scattering is subtracted from this to obtain the $I(s)$ for the protein in the solution, which can then be analysed (Mertens and Svergun, 2010).

1.6.2.3 Analysis of SAXS data

SAXS experiments involve exposing the sample and reference buffer to X-rays, the scattering of which is collected with a specific SAXS detector (Fig. 1-18). Following data collection, primary analysis of the scattering data yields parameters that can then be used to obtain *ab initio* models of the studied protein. The ATSAS software developed by the Biological Small Angle Scattering Group lead by Dimitri Svergun is an advanced programme package that can be used to analyse all aspects of SAXS data (Petoukhov *et al.*, 2012). PRIMUS is used for initial data reduction and processing such as buffer subtraction to obtain the final scattering curve (Fig. 1-20A & C). Guinier analysis can be carried out to determine forward scattering intensity at zero scattering angle, $I(0)$, and the radius of gyration, R_g (Mertens and Svergun, 2010). The Guinier-plot is a good way of evaluating sample quality. A non-linear plot with increasing intensity at low angles is indicative of aggregation while decreasing intensity at low angles suggests inter-particle repulsion (Fig. 1-20B) (Mertens and Svergun, 2010).

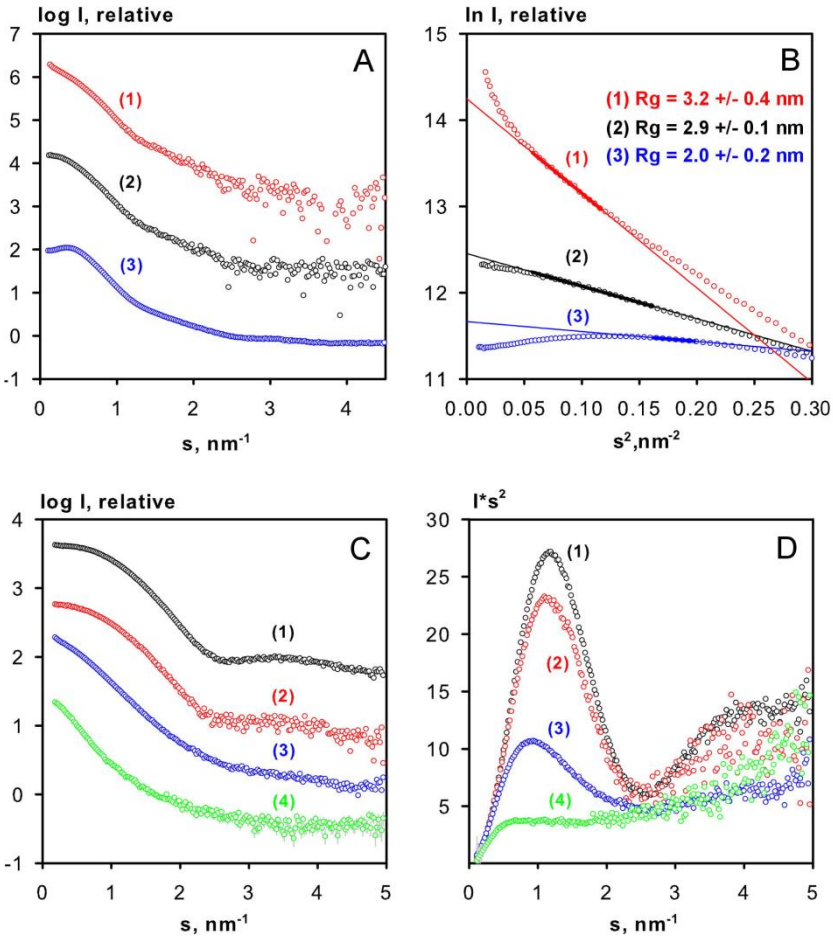


Figure 1-20. Examples of SAXS data curves
The figure shows (A) scattering curves and (B) Guinier plots for samples of BSA. The data in red (1) are an example of an aggregated sample, in black (2) an ideal sample resulting in good data and in blue (3) a sample showing inter-particle interaction. (C) The scattering curve and (D) the Kratky plot for lysozyme. The data in black (1) are from folded lysozyme, in red (2) partially unfolded lysozyme (in 8 M urea), data in blue (3) are also from partially unfolded lysozyme (at 90°C) and in green (4) the lysozyme is completely unfolded (in 8 M urea at 90°C). This figure was taken from Mertens and Svergun (2010) with permission from Elsevier.

The molecular mass (M) of the studied protein can be determined using its forward scattering intensity, $I(0)$, and that of a control protein with a known molecular mass, such as BSA

$$M \approx M_{BSA} * \frac{[I(0)/c]_{protein}}{[I(0)/c]_{BSA}} \qquad (1.24)$$

PRIMUS can also be used to obtain the mass of a protein from the hydrated particle volume, or the Porod volume, which is 1.5-2 times the molecular mass in kDa (Mertens and Svergun, 2010). The folding state of the protein of interest can be determined with the Kratky plot (Fig. 1-19D, Doniach, 2001). Globular folded proteins will have a bell shaped

curve at low angles. With unfolded proteins the curve continuously increases and partially folded proteins or muti-domain proteins with linkers give curves between these two.

The programme GNOM (Svergun, 1992) can be used to obtain the maximum dimension parameter, D_{max}, of a protein (Fig. 1-21). This is also an alternative method of obtaining the $I(0)$ and R_g in real space (Mertens and Svergun, 2010). These parameters are determined from the distance distribution function, $p(r)$, where r is the distance between pairs of scattering electrons within a molecule (Serdyuk *et al.*, 2007c). The $p(r)$ is an indirect Fourier transformation of the scattering intensity (Mertens and Svergun, 2010)

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty s^2 I(s) \frac{\sin sr}{sr} ds \qquad (1.25)$$

where the integral over 0 - ∞ is the D_{max} value (Serdyuk *et al.*, 2007c). The output file generated by GNOM containing the above information can then be used for *ab initio* modelling.

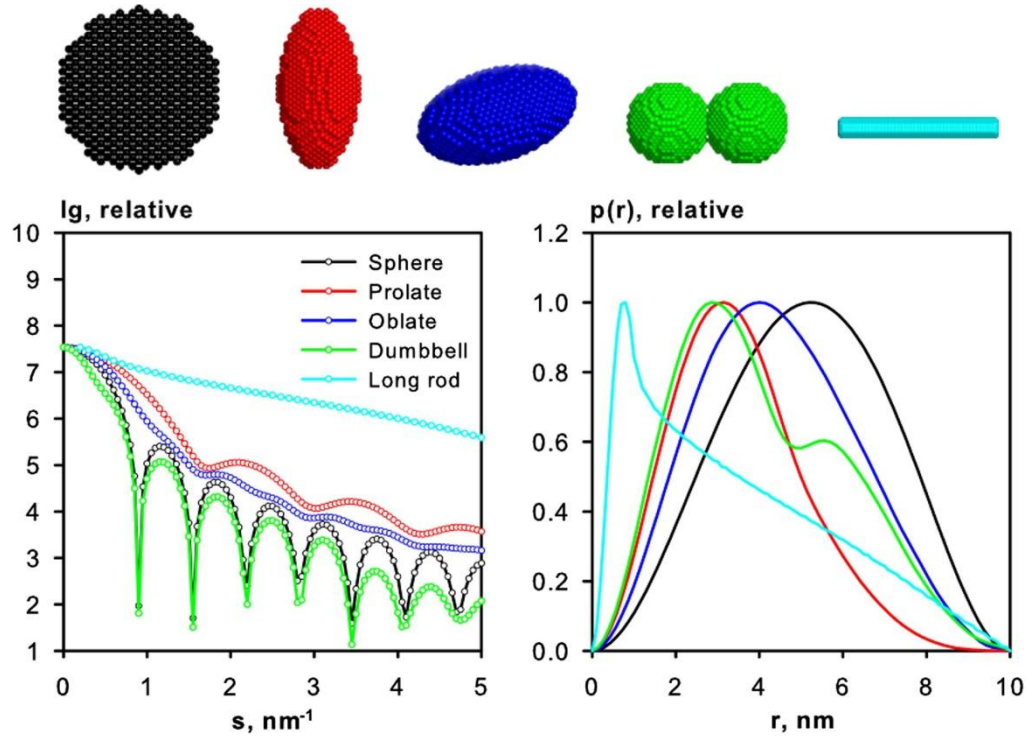


Figure 1-21. The scattering curve and p(r) profiles for various molecules
The scattering curve and distance distribution function for several different shapes calculated from bead models shown above the plots. This figure was taken from Mertens and Svergun (2010) with permission from Elsevier.

1.6.2.4 Ab initio modelling of SAXS data

Ab initio modelling of proteins in this study was carried out using DAMMIF (Franke and Svergun, 2009). The programme uses spheres, or dummy atoms, to model the scattering envelope using a simulated annealing method as described for the original widely used programme DAMMIN (Svergun, 1999). DAMMIN fills a search volume (a sphere with a radius (R) equal to half the maximum particle size (D_{max}) with densely packed spheres with a radius of $r_0 << R$, which are either assigned to the particle or solvent. A random dummy atom model (DAM) is formed and the scattering curve for that model is calculated and compared with the experimental data. To form a DAM representative of the experimental particle, a single solvent dummy atom is then chosen at random and assigned to be part of the particle. The scattering curve of the new DAM is calculated, compared with the experimental scattering data and penalties are implemented to take into account that for the final model all the dummy atoms must be interconnected. If the chosen sphere makes the DAM disconnected, the dummy atom is rejected from the model, likewise if connectivity is maintained the dummy atom is kept in the model. This is continued one sphere at a time until a satisfactory fit to the experimental data is reached. Calculating the scattering amplitudes for each potential model before accepting or rejecting the change made is very CPU intensive, thus DAMMIF (the F denoting *fast*) was developed, which is 25-40 times faster than DAMMIN (Franke and Svergun, 2009). DAMMIF calculates the connectivity of a randomly selected sphere to the particle first before moving forward to calculate the scattering amplitudes. Also while DAMMIN completely randomises which beads are in particle and solvent phase within the search volume, DAMMIF starts with an isometric particle with a radius of gyration (R_g) determined from the experimental data, which reduces the amount of intermediate DAMs that need to be processed. In addition, DAMMIF was programmed to include an unlimited search volume. In DAMMIN, the initial search volume is defined before starting and this cannot be changed while the programme is running. This can lead to artefacts in cases where a sphere chosen for the particle phase is close to the border of the search volume. As the programme will not allow the DAM to extend past the borders, effects like artificial bending of the models can occur. DAMMIF has taken this into account by not limiting the search volume. If a dummy atom assigned to the particle is near the border, a new solvent phase bead is added to allow expansion and to make sure that solvent beads always surround the particle beads. Throughout modelling the structural similarity of the generated models is monitored using

normalised spatial discrepancy (NSD) and the model R_g. Higher NSD values indicate higher model flexibility.

Ab initio modelling of flexible proteins cannot be done with DAMMIF, which is appropriate for macromolecules with more rigid structures. Flexible proteins can be modelled using the Ensemble Optimisation Method (EOM) (Bernado *et al.*, 2007; Petoukhov *et al.*, 2012). The programme can be used to model intrinsically unfolded proteins or multi-domain proteins with flexible linkers. The method is based on the assumption that the flexible protein can be in an average number of conformations, which represent the experimental data. The programme first generates a pool containing a large number of conformations (the default is 10,000), which cover the experimental configurational space. From this, a genetic algorithm is used to assign the appropriate models to subsets to generate an ensemble of models that fit the experimental data. The D_{max} and R_g of the ensemble are determined and these can then be compared with the respective values obtained for the pool. The more variation there is in the D_{max} and R_g values of the ensemble, the more flexible a protein is. Likewise, if the ensemble values lie within a tight peak compared with the pool, the protein is rigid. Representative structural envelope models are also generated for the ensemble. However, care must be taken not to over interpret the models as they are based on given crystal structures or homology models and, due to the flexibility of a protein, any of the given models may be as correct as another.

Chapter 2

Materials and methods

2.1 Chemicals, reagents and kits

Agilent Technologies	StrataClone PCR Cloning Kit
Amersham	Autoradiography film
Applied Biosystems	POROS® PEEK™ 4.6 mm x 100 mm 1.7 ml column, POROS® MC 20 µm Self Pack® Media
BD Biosciences	Mouse monoclonal anti-His-antibody
BDH	Saponin
Bio-Rad	Bradford protein assay reagent, Econo-Pac® chromatography columns, Gene pulser electroporation cuvette (0.2 cm), Precision plus dual colour protein standard
Blood Transfusion Service	Human full blood
BOC	Malaria culture gas (5% CO_2, 1% O_2, 94% N_2)
Eurogentec	Oligonucleotides
Fisher Scientific	10x Phosphate buffered saline (PBS), ammonium persulphate, glycerol, low absorbance imizadole, Tris-base
GE Healthcare	Amersham Hybond-N+ nylon membrane, CDP Star detection solution, Gene Images AlkPhos Direct Labelling kit, GSTrap FF 1 ml column, HiLoad™ 16/60 Superdex 200 gel filtration column, HiPrep 26/60 Sephacryl S-400 HR gel filtration column, HisTrap FF 1 ml column, Hyperfilm ECL, PD-10 desalting columns, PreScission Protease, Vivaspin™ sample concentrators
GENEART AG*	*E. coli* codon optimised gene synthesis (*from 2010 onwards: GeneArt® Gene Synthesis, Life Technologies)
Jacobus Pharmaceuticals	WR99210
Kodak	X-ray film
Life Technologies	Accuprime™ *Pfx* SuperMix, SybrSafe, Albumax, NuPAGE Bis/Tris 10% polyacrylamide gels, NuPAGE® MES SDS Running Buffer, Ultrapure™ agarose, Ultrapure™ low melting point agarose

Melford	1,10-phenanthroline, ampicillin, benzamidine, dithiothreitol (DTT), E-64, isopropyl β-D-thio-galactopyranoside (IPTG), leupeptin, β-nicotinamide adenine dinucleotide (NAD^+), reduced β-nicotinamide adenine dinucleotide (NADH), proteinase K
Millipore	Aquacide II, Immobilon western chemiluminescent HRP substrate, Millex®-HV PVDF 0.45 μm syringe filters
Nalgene	1.0 ml cryotubes
New England Biolabs	1 kb DNA ladder, all restriction enzymes and their buffers, blue gel loading dye (6x), T4 DNA ligase
Novagen	BugBuster™ protein extraction reagent, benzonase
PAA The Cell Culture Company	Gentamycin, RPMI 1640 (with 25 mM HEPES and L-glutamine)
Promega	pGEM-T-Easy Vector System I
Qiagen	Qiaquick gel extraction kit, Qiaquick PCR purification kit, Qiaprep spin DNA maxiprep kit, Qiaprep spin DNA miniprep kit, Ni-NTA agarose
Roche	DNase
Sigma	5-bromo-4-chloro-3-inodlyl-β-D-galactopyranoside (X-Gal), acrylamide/bisacrylamide 29:1, alcohol dehydrogenase, apoferritin, bovine serum albumin (BSA), butanol, chloramphenicol, coenzyme A, dextran blue, dihydrolipoamide, D-sorbitol, flavin adenine dinucleotide (FAD), glycine, hypoxanthine, kanamycin sulphate, lipoamide, lithium potassium acetyl phosphate, lysozyme, pepstatin A, phenol:chloroform:isoamyl alcohol 25:24:1, phenylmethanesulphonyl fluoride, phosphotransacetylase, Ponceau S, N,N,N',N'-tetramethylethylenediamine (TEMED), thryoglobulin
Takara	TaKaRa Ex Taq™ polymerase (+ Mg^{2+} free buffer)
Thermo Scientific	GelCode Blue Safe Stain, Restore western blot stripping Buffer, SnakeSkin® 10 K MWCO 16 mm dry dialysis tubing
VWR	Giemsa stain
Whatman	Protan® nitrocellulose membrane

2.2 Equipment

Beckman Coulter	Avanti J-26 XP centrifuge with rotors JA25.50 and JSP F10, Optima analytical ultracentrifuge XL-I with An-60 Ti 4 and An-50 Ti 8 hole rotors, AUC sample cell parts: 12 mm Epon charcoal-filled double sector centrepieces, sapphire windows, gaskets, screws and housing
Bio-Rad	Agarose gel tanks, GelDoc 2000, Trans-Blot® SD Semi-Dry transfer cell
Clare Chemical Research	DarkReader transilluminator
Constant Systems	One shot cell disrupter
Eppendorf	Microcentrifuge 5415D with rotor F45-24-11, microcentrifuge 5415R with rotor F45-24-11
Fisher Scientific	AccuSpin™ MicroR refrigerated microcentrifuge with biocontainment rotor 13-100-515, agarose gel tanks
GE Healthcare	ÄKTApurifier, UNICORN 5.1 software
G Storm	PCR machine
Grant	Heat block QBT2
Heidolph	Titramax 100 shaker
IKA	VXR basic Vibrax® with test tube insert
Kuehner	Shaking incubator ISF-1-W
Life Technologies	Xcell Surelock™ Mini-cell apparatus
Perkin Elmer	Wallac 1450 MicroBeta Trilux liquid scintillation & luminescence counter, printed filter mat A
Sanyo	CO_2 incubator MCO-18A1C
Scientific Industries	Genie 2 Vortex mixer
Sciquip	Sigma 6K 15 centrifuge with rotors 11150 and 12500
Shimadzu	UV-2501 spectrophotometer
Thermo Scientific	NanoDrop 1000 spectrophotometer, Heraeus Megafuge 16R with rotor 750033629
Tom Tech	Harvester 96™ Mach III
UVP	HS-1000 hybridization oven and UV crosslinker
Wolf Laboratories	Galaxy S CO_2 incubator

2.3 Buffers, solutions, media

2.3.1 DNA analysis

1x TAE	40 mM Tris-acetate, 1 mM EDTA, pH 8.0
6x DNA loading dye	0.25% (w/v) bromophenol blue, 0.25% (w/v) orange-G, 0.25% (w/v) xylene cyanol FF, 40% (w/v) sucrose
20x SSC	300 mM trisodium citrate, 3 M NaCl, pH 7-8
Colony cracking lysis buffer	20% (w/v) sucrose, 0.5 M NaOH, 0.3 M KCl, 0.025 M EDTA, 1.25% SDS
Southern blot hybridisation buffer	Gene images AlkPhos hybridization buffer, 0.5 M NaCl, 4% (w/v) blocking reagent
Southern primary wash buffer	2 M Urea, 150 mM NaCl, 1 mM $MgCl_2$, 50 mM sodium phosphate pH 7.0, 0.1% (w/v) SDS, 0.2% (w/v) blocking reagent
Southern secondary wash buffer	50 mM Tris-base, 100 mM NaCl, 2 mM $MgCl_2$
TE	10 mM Tris-HCl, 1 mM EDTA, pH 8.0
X-Gal stock solution	2% (w/v) 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) in dimethylformamide

2.3.2 Protein analysis

6x SDS loading dye	62.5 mM Tris-HCl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 0.001% (w/v) bromophenol blue, 5% (v/v) mercaptoethanol
Resolving gel	375 mM Tris-HCl, pH 8.9, 0.1% (w/v) SDS, 10% (v/v) acrylamide
Stacking gel	122 mM Tris-HCl, pH 6.7, 0.1% (w/v) SDS, 5% (v/v) acrylamide
Towbin buffer	25 mM Tris, 192 mM glycine, 20% (v/v) methanol pH 8.3
Tris/glycine resolving buffer	25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3

2.3.3 Protein purification

Table 2-1. Buffers used for purification of recombinant proteins

Purification method	Buffer name	Buffer composition
HisTrap FF	HisTrap binding buffer	20 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 7.4
	HisTrap buffer A	20 mM sodium phosphate, 300 mM NaCl, pH 7.4
	HisTrap buffer B1 (His-r*Pf*E2m)	20 mM sodium phosphate, 300 mM NaCl, 1 M imidazole, pH 7.4
	HisTrap buffer B2 (His-r*Pf*E2bc)	20 mM sodium phosphate, 150 mM NaCl, 2 M imidazole, pH 7.4
POROS	POROS binding buffer	50 mM potassium phosphate, 150 mM NaCl, 0.1 mM EDTA, 10 mM imidazole, pH 8.0
	POROS buffer A	50 mM potassium phosphate, 150 mM NaCl, 0.1 mM EDTA, pH 8.0
	POROS buffer B	50 mM potassium phosphate, 150 mM NaCl, 0.1 mM EDTA, 1 M imidazole, pH 8.0
Ni-NTA	Qiagen lysis buffer	50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 8.0
	Qiagen wash buffer 1	50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 8.0
	Qiagen wash buffer 2	50 mM sodium phosphate, 300 mM NaCl, 50 mM imidazole, pH 8.0
	Qiagen elution buffer	50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0
GSTrap	GSTrap binding buffer	1x PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3)
	GSTrap elution buffer	50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0
	PreScission cleavage buffer	50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, pH 7.0
Gel filtration	PEBS150	50 mM potassium phosphate, 150 mM NaCl, 2 mM EDTA, pH 7.4
	PEBS150 + glycerol + DTT	50 mM potassium phosphate, 250 mM NaCl, 2 mM EDTA, 0.1% glycerol, 5 mM DTT, pH 7.4
	PEBS250	50 mM potassium phosphate, 250 mM NaCl, 2 mM EDTA, pH 7.4
Column maintenance	Stripping buffer	1 M NaCl, 50 mM EDTA, pH 8.0

2.3.4 Bacterial culture

Ampicillin	100 mg/ml in dH_2O
Chloramphenicol	34 mg/ml in 100% ethanol
Kanamycin	50 mg/ml in dH_2O
Luria Bertani (LB) medium	10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl
LB agar	15 g/L agar into LB medium
Super optimal broth with catabolite repression (SOC) medium	20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl, 10 mM $MgSO_4$, 20 mM glucose
Terrific Broth (TB) medium	12 g/L tryptone, 24 g/L yeast extract, 0.4% glycerol, 9.4 g/L K_2HPO_4, 2.2 g/L KH_2PO_4, pH 7.2
Tfb I	100 mM RbCl, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2$, 5% (v/v) glycerol
Tfb II	10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$, 15% (v/v) glycerol

2.3.5 Parasite culture

5-fluorocytosine (5-FC)	4 mM 5-FC in dH_2O
Blood wash medium	16.37 g/L RPMI 1640 (with 25 mM HEPES and L-glutamine), 0.1% (w/v) sodium bicarbonate, 10 mM glucose, 20 μM hypoxanthine, 20 μg/ml gentamycin, pH 7.4
Complete RPMI 1640 medium	Blood wash medium with 0.5% (w/v) Albumax II
Cytomix	120 mM KCl, 0.15 mM $CaCl_2$, 2 mM EGTA, 5 mM $MgCl_2$, 10 mM potassium phosphate, pH 7.6, 25 mM HEPES, pH 7.6
IC_{50} medium	Complete RPMI 1640 medium without hypoxanthine
Freezing solution	PBS with 30% (v/v) glycerol
Sorbitol solution	5% (w/v) sorbitol in 10 mM potassium phosphate, pH 7.2

Thawing solution 1	12% (w/v) NaCl
Thawing solution 2	1.8% (w/v) NaCl
Thawing solution 3	0.9% (w/v) sodium chloride, 0.2% (w/v) glucose
WR99210	Stock solution: 20 mM in dimethylsulphoxide Working solution: 20 μM in blood wash medium

2.4 Vectors, primers, constructs, bacterial and *P. falciparum* strains, antibodies

2.4.1 Vectors

Vectors used in this study were for cloning, protein expression and *P. falciparum* gene knockout.

Table 2-2. Vectors and their selectable markers, use and source

Vector	Selectable marker	Use	Source
pGEM-T-Easy	ampicillin	cloning	Promega
pSC-B-amp/kan	ampicillin/kanamycin	cloning	Agilent Technologies
pET-15b (N-terminal His-tag)	ampicillin	protein expression	GENEART AG
pGEX-6P-1 (N-terminal GST-tag)	ampicillin	protein expression	GE Healthcare
pCC-1	Ampicillin/human dehydrofolate reductase (*hdhfr*)/*Saccharomyces cerevisiae* cytosine deaminase and uracil phosphoribosyl transferase chimera	*P. falciparum* gene knockout	Maier *et al*., 2006

2.4.2 Primers

Table 2-3. Primers used in this study

Construct	Name of primer	Direction	Primer sequence
His-r*Pf*E2bc (pET-15b)	*Pf*E2 binding domain	fwd	GCGC CATATG GAA ATG GTT CTG CCG AGC (*Nde*I)
	*Pf*E2 catalytic domain	rev	GCGC GGATCC TTA TTA CAG AAA GAT CTT CAT GAT GTC C (*Bam*HI)
GST-r*Pf*E2c (pGEX-6P-1)	*Pf*E2 catalytic domain	fwd	GCGC GGATCC AAA ACC GTT GAT ATG ACC AAT ATC C (*Bam*HI)
	*Pf*E2 catalytic domain	rev	GCGC CTCGAG TTA TTA CAG AAA GAT CTT CAT GAT GTC C (*Xho*I)
pCC1 *Pf*E2 knockout construct	pCC Δ*Pf*E2 5'	fwd	GAGC CCGCGG AGT TTT AAC TAC AGA GGA AAA CGA AAA TAT G (*Sac*II)
	pCC Δ*Pf*E2 5'	rev	GAGC CTTAAG GGC AAC TTC TTT GTC TAA ATC TAC GAA C (*Afl*II)
	pCC Δ*Pf*E2 3'	fwd	GAGC GAATTC CCT TTT ATA GAA ACA GGA AAA CAC CC (*Eco*RI)
	pCC Δ*Pf*E2 3'	rev	GAGC CCATGG CGC ATT TTA CAA ACC TTC CAT TC (*Nco*I)
pCC1 *Pf*aE3 knockout construct	pCC Δ*Pf*aE3 5'	fwd	GAGC ACTAGT CTT AAC GTC GTT ACT CTA ATT TGG TTA TC (*Spe*I)
	pCC Δ*Pf*aE3 5'	rev	GAGC CTTAAG GCG CTT TGC TTG GTA TAC AGC C (*Afl*II)
	pCC Δ*Pf*aE3 3'	fwd	GAGC GAATTC GCA CAC ACA GCA TCA TAT CAA G (*Eco*RI)
	pCC Δ*Pf*aE3 3'	rev	GAGC CCTAGG TTA GTG AGT TCT TAT TTT TGA TAT AGA (*Avr*II)

2.4.3 Constructs

Table 2-4. The insert, vector and purpose of the constructs used in this study

Name	Insert	Vector	Purpose
His-r*Pf*E2m	*E. coli* codon optimised mature length *Pf*E2, amino acids 51 - 640	pET-15b	protein expression
His-r*Pf*E2bc	*E. coli* codon optimised binding and catalytic domain of *Pf*E2, amino acids 356 - 640	pET-15b	protein expression
GST-r*Pf*E2c	*E. coli* codon optimised catalytic domain of *Pf*E2, amino acids 429 - 640	pGEX-6P-1	protein expression
His-r*Pf*aE3-pET-15b	*E. coli* codon optimised mature length apicoplast *Pf*aE3, amino acids 101 - 667	pET-15b	protein expression
His-r*Pf*aE3	Native mature *Pf*aE3, amino acids 111 - 667	pQE30	protein expression
pCC1 Δ*Pf*E2	5' flank (nucleotides 685-1190) and 3 ' flank (nucleotides 2119-2626) of *Pf*E2	pCC1	knockout of *PfE2*
pCC1 Δ*Pf*aE3	5' flank (nucleotides 37-526) and 3 ' flank (nucleotides 1456-2001) of *Pf*aE3	pCC1	knockout of *PfaE3*

2.4.4 Chemically competent bacterial strains

Table 2-5. The genotype, use and source of chemically competent cells used in this study

Chemically competent cells	Genotype	Use	Source
DH5α	F$^-$ Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (r_K^-, m_K^+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	cloning	Life Technologies
StrataClone SoloPack Competent Cells	F$^-$ Φ80*lac*ZΔM15 *rec*A1 *end*A1 (r_K^-) *tonA* (Strep^R)	cloning	Promega
BLR(DE3)	F$^-$ *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm* (DE3) Δ(*srl-rec*A)306::Tn10 (Tet^R)	protein expression	Novagen
Rosetta™ 2(DE3)pLysS	F$^-$ *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm* (DE3) pLysSRARE* (Cam^R)	protein expression	Merck Millipore
BL21 Star™(DE3)	F$^-$ *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm* (DE3) *rne131*	protein expression	Life Technologies
NovaBlue(DE3)	*end*A1 *hsd*R17(r_K^- m_K^+) *supE44 thi-1 recA1 gyrA96 relA1 lac* (DE3) F'[*proA*$^+$*B*$^+$ *lacI Z*ΔM15::Tn10] (Tet^R)	protein expression	Novagen

*contains tRNA genes *argU*, *argW*, *ileX*, *glyT*, *leuW*, *proL*, *metT*, *thrT*, *tyrU*, and *thrU*. Rare codons AGG, AGA, AUA, CUA, CCC, and GGA are supplemented.

2.4.5 *P. falciparum* strains

D10	Papua New Guinea
3D7	The Netherlands

2.4.6 Antibodies

Antibodies in this thesis were used for western blot analyses.

Table 2-6. Antibodies and their dilutions and sources

Antibody	Raised in	Dilution for western blot	Source
Primary antibodies			
α-His	mouse	1:25000	BD Biosciences
α-GST	rabbit	1:50000	GE Healthcare
α-LA	rabbit	1:10000	Eurogentec
Secondary antibodies			
α-mouse IgG (H+L) HRP conjugate	N/A	1:10000	Promega
α-rabbit IgG (H+L) HRP conjugate	N/A	1:10000	Promega

2.5 Molecular biology methods

2.5.1 Polymerase Chain Reaction (PCR)

2.5.1.1 PCR of inserts for recombinant protein expression constructs

For production of the His-r*Pf*E2bc and GST-r*Pf*E2c inserts (Table 2-4), 1 µl of each primer (final concentration 0.4 µM) and 1 µl of the template were added to 22 µl of AccuPrime™ Pfx SuperMix. The template used for the reaction was *E. coli* codon optimized His-r*Pf*E2m construct (Table 2-4). The products from this reaction are blunt ended.

The PCR conditions used were:

Step 1	94°C, 4 min
Step 2	94°C, 30 s
Step 3	55 – 45°C, 30 s (touchdown method)
Step 4	68°C, 1 min
Repeat steps 2 - 4 for 35 cycles	
Step 5	68°C, 10 min
Step 6	storage at 10°C

2.5.1.2 PCR of inserts for *P. falciparum* gene knockout

For double-crossover gene knockout, ~500 bp fragments homologous to the 5' and 3' ends of *Pf*E2 and apicoplast *Pf*E3 (*Pf*aE3) were generated. For the *Pf*E2 5' and 3' flanks and *Pf*aE3 5' flank (Table 2-4) AccuPrime™ Pfx SuperMix was used as described above. *P. falciparum* genomic DNA was used as template.

The PCR conditions were:

Step 1	94°C, 4 min
Step 2	94°C, 30 s
Step 3	48.5°C, 30 s
Setp 4	68°C, 1 min
Repeat steps 2 – 4 for 35 cycles	
Step 5	68°C, 10 min
Step 6	storage at 10°C

The *Pf*aE3 3' flank (Table 2-4) was amplified using TaKaRa Ex Taq™ Polymerase. The 50 µl reaction mix consisted of 1.25 U of TaKaRa Ex Taq™ Polymerase, 0.2 mM of each dNTP, <500 ng genomic DNA template, 0.3 µM of each primer in 1x Ex Taq™ buffer. The products from this PCR method contain TA overhangs.

The PCR conditions were:

Step 1	94°C, 4 min
Step 2	94°C, 30 s
Step 3	52°C, 1 min
Setp 4	68°C, 1 min
Repeat steps 2 – 4 for 35 cycles	
Step 5	68°C, 10 min
Step 6	storage at 10°C

2.5.2 Cloning

2.5.2.1 Cloning recombinant protein expression constructs into intermediate vector

The PCR products for His-r*Pf*E2bc and GST-r*Pf*E2c constructs, generated in section 2.5.1.1, were purified using the Qiaquick PCR purification kit, as per manuafacturers instructions. The blunt ended fragments were then cloned into the intermediate vector

pSC-B-amp/kan (Fig. 2-1) using the StrataClone Blunt PCR Cloning Kit. The StrataClone Blunt Vector Mix contains two blunt-ended vector arms with covalently bound topoisomerase I on one end and a *loxP* recognition site on the other. The PCR fragment is ligated into the vector arms via the topoisomerase I. This intermediate is transformed into competent bacteria that transiently express Cre recombinase, which then mediates the recombination of the two *loxP* sites to form a circular plasmid. The ligation mixture contained the following components in order of addition: 3 µl StrataClone Blunt Cloning Buffer, 2 µl PCR product (5-50 ng) and 1 µl StrataClone Blunt Vector Mix. The ligation mix was incubated at room temperature overnight and then stored at 4°C until transformation in the afternoon.

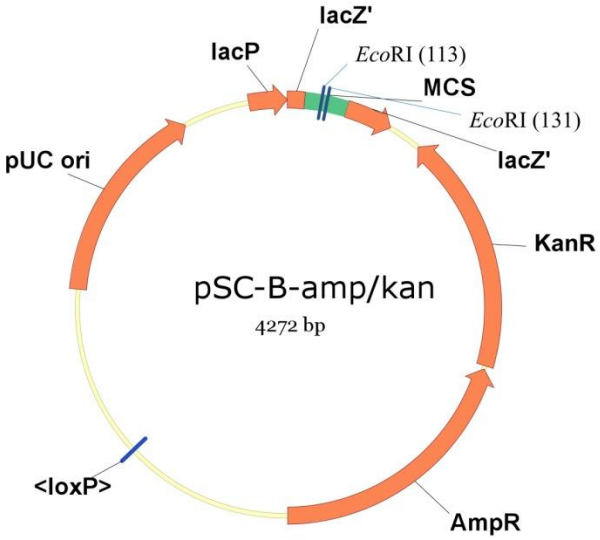


Figure 2-1. Vector map of pSC-B-amp/kan
The pSC-B-amp/kan vector contains amipcillin (AmpR 1268-2125 bp) and kanamycin (KanR 465-1256 bp) resistance cassettes. The multiple cloning site (MCS 57-197 bp, green) contains the PCR product insertion site at base pair 123, which is flanked by *Eco*RI restriction sites, which can be used to perform diagnostic digests after ligation. The multiple cloning site (MCS) is in the middle of the β-galactosidase α-fragment coding sequence (lac*Z*' 1-354 bp), which allows blue-white colony screening in the presence of X-Gal. The lac*Z*' is under the control of the *lac* promoter (lacP 4153-4272 bp). The pUC origin of replication (pUC ori 3264-3931 bp) is present. <loxp> (2690-2723 bp) denotes where the two *loxP* recognition sites have been recombined to form the circular plasmid.

StrataClone SoloPack competent cells (Table 2-5) were transformed as described in section 2.5.4. As the template used for the PCR of His-r*Pf*E2bc and GST-r*Pf*E2c inserts was a plasmid containing an ampicillin (Amp) resistance cassette, the transformation mix was grown on plates containing kanamycin (Kan) to prevent any carryover of template plasmid.

Blue-white colour screening was used to select colonies positive for the plasmid containing an insert. A total of 40 µl of 2% (v/v) X-Gal (section 2.3.1) was spread onto Luria Bertani/kanamycin (LB/Kan) agar plates and allowed to dry before spreading the ligation mix. Colonies containing the plasmid (white colonies) were then screened using the colony cracking method (section 2.5.5). This was followed by plasmid mini preparations from selected colonies (section 2.5.6) and a restriction digest analysis (section 2.5.7) with *Eco*RI (Fig. 2-1). The sequence of plasmids displaying the correct restriction digest profile was determined (Source BioScience, LifeSciences) before the insert was subcloned into the final destination vector.

2.5.2.2 Cloning recombinant protein expression constructs into final vectors

The His-r*Pf*E2bc and GST-r*Pf*E2c inserts were excised from the intermediate pSC-B-amp/kan vector using the appropriate restriction enzymes (*Nde*I/*Bam*HI and *Bam*HI/*Xho*I, respectively). The final vectors pET-15b (Fig 2-2) and pGEX-6P-1 (Fig 2-3), respectively, were digested with the same enzymes to allow directional cloning of the inserts.

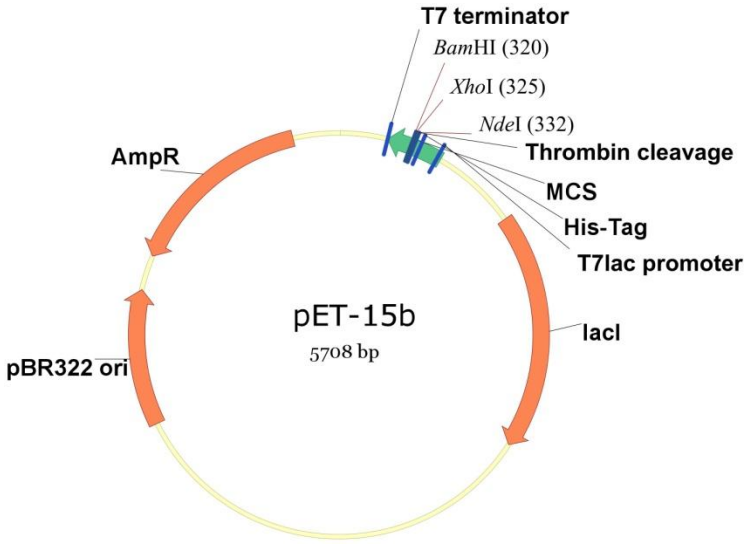


Figure 2-2. Vector map of pET-15b
The pET-15b cloning/expression region (green) contains an N-terminal His-Tag™ (362-380 bp) followed by a thrombin cleavage site (336-353 bp) and the MCS (319-335 bp) with *Nde*I, *Xho*I and *Bam*HI restriction sites. Expression is controlled by the T7*lac* promoter (436-469 bp), which in turn is repressed by *lacI* (LacI 866-1945 bp) for induction of protein expression with IPTG. The T7 terminator (213-259 bp) is found after the MCS. The plasmid contains the pBR322 origin of replication (pBR322 3882-4501 bp) and has an ampicillin resistance cassette (AmpR 4643-5500 bp).

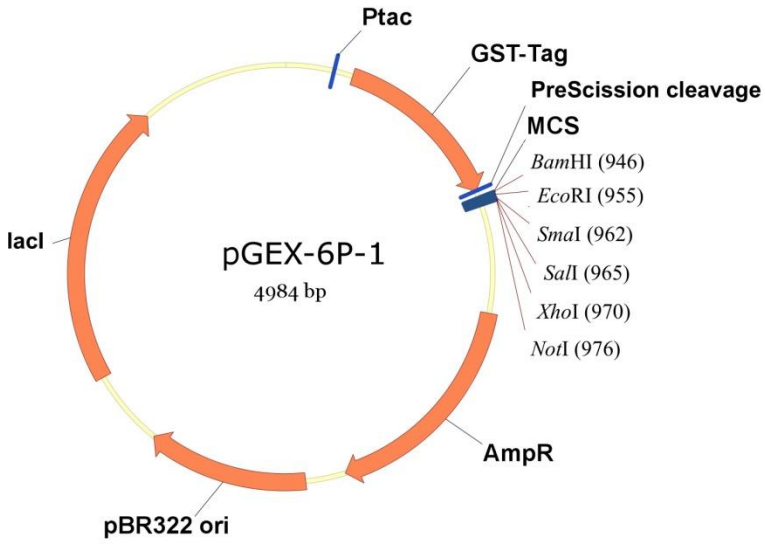


Figure 2-3. Vector map of pGEX-6P-1
The pGEX-6P-1 vector has an N-terminal GST-Tag™ (258-992 bp) followed by a PreScission™ Protease cleavage site (PreScission 918 - 938bp) and the MCS (943-979 bp, blue) with *Bam*HI, *Eco*RI, *Sma*I, *Sal*I, *Xho*I and *Not*I restriction sites. Expression is controlled by the tac promoter (Ptac 184-212 bp), which in turn is repressed by *lacI* (LacI 3324-4415 bp). The plasmid contains the pBR322 origin of replication (pBR322 2407-3026 bp) and has an ampicillin resistance cassette (AmpR 1392-2252 bp).

Digestion reactions (50 µl) contained 9 µl of either insert of vector (~ 3 µg), 1 µl of each restriction enzyme, 5 µl of the appropriate NEB buffer and 34 µl of dH_2O. The mix was incubated for 4-5 h at 37°C after which the digested DNA was separated on 1 % (w/v) Ultrapure™ low melting point agarose. The insert and vector were excised and extracted from the gel as described in section 2.5.9.

Ligation of the inserts into the vectors was carried out using T4 DNA ligase (New England Biolabs). Two ligations for each vector/insert pair were prepared with varying amounts of insert. The ligation mix contained 3 µl of vector, 6 or 12 µl of insert, 1 µl of T4 DNA ligase, 2 µl of 10 x T4 DNA ligase buffer and was made up to 20 µl with dH_2O. Ligation was carried out overnight at room temperature. In the morning the mixtures were transferred into storage at 4°C and transformation into DH5α chemically competent cells (Table 2-5) was carried out in the afternoon as described in section 2.5.4 using LB/ampicillin (LB/Amp) agar plates. Restriction digests (section 2.5.7) were carried out to determine transformed colonies, and plasmid with the correct restriction digest profiles was

sequenced (Source BioScience, LifeSciences) before subsequent use in recombinant protein expression.

2.5.2.3 Cloning *Pf*E2 and *Pf*aE3 knockout constructs into intermediate vectors

The *Pf*E2 5' and 3' flanks and *Pf*aE3 5' flanks were cloned into the pSC-B-amp/kan intermediate vector as above (section 2.5.2.1). PCR of *Pf*aE3 3' was carried out using TaKaRa Ex Taq™ Polymerase (section 2.5.1.2), which creates TA overhangs; therefore the PCR fragments were subcloned into the pGEM®-T Easy Vector (Fig. 2-4) using the pGEM®-T Easy Vector System kit (Promega).

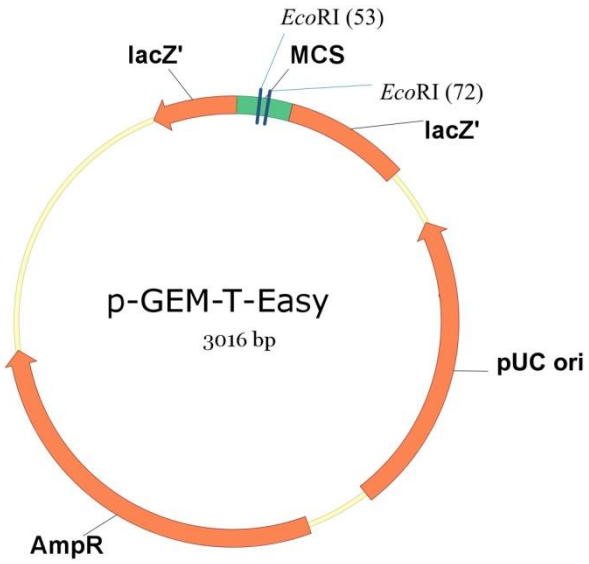


Figure 2-4. Vector map of pGEM®-T Easy Vector
The MCS (10-128 bp) for the pGEM®-T Easy Vector is located within the β-galactosidase α-fragment coding sequence (lacZ' 166-395, 2836-2996 bp). The MCS is flanked by *Eco*RI restriction sites, which can be used to cut the PCR fragment out with a single digest. The vector has an ampicillin resistance cassette (AmpR 1337-2197 bp) and a pUC origin (523-1190 bp).

The ligation reaction (10 µl) contained 5 µl 2x Rapid Ligation Buffer, 1 µl pGEM®-T Easy Vector, ~80 ng purified PCR product and 1 µl T4 DNA ligase. The mixture was incubated overnight at room temperature. In the morning the ligations were transferred into storage at 4°C and transformation into DH5α chemically competent cells (Table 2-5) was carried out in the afternoon. Blue/white selection was carried out on LB/Amp agar plates

supplemented with 0.5 mM isopropyl β-D-thio-galactopyranoside (IPTG) and 80 µg/ml X-Gal. Colony cracking (section 2.5.5) was carried out to screen colonies positive for plasmid, which was followed by restriction digest analysis (section 2.5.7) to determine which plasmid DNA to send for sequencing (The Sequencing Service, Dundee) before subcloning into final vector.

2.5.2.4 Cloning *Pf*E2 and *Pf*aE3 knockout constructs into final vector

Ligation of each flank into the final pCC-1 vector (Fig. 2-5) was carried out one at a time as described in section 2.5.2.2. For restriction enzymes used for each flank see Table 2-3. The 5' flank was cloned first into the pCC-1 vector. This was then sequenced (The Sequencing Service, Dundee) before inserting the 3' flank into the same vector. A final sequencing of the finished construct was carried out before transfection into *P. falciparum.*

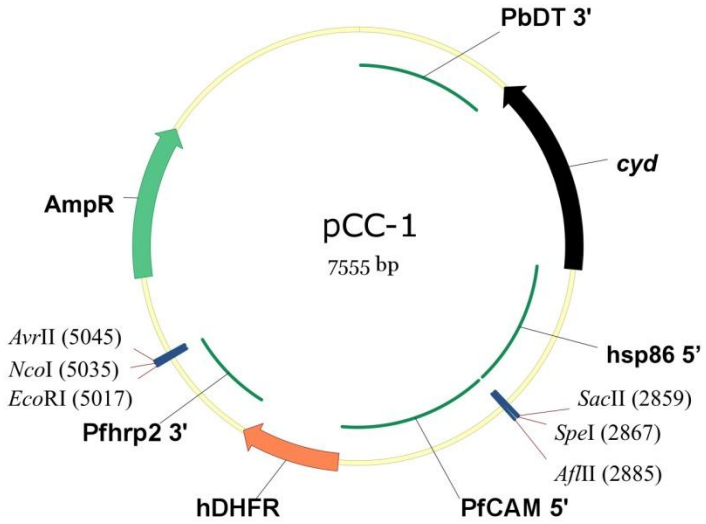


Figure 2-5. Vector map of pCC-1
The pCC-1 vector contains multiple cloning sites at 2856-2888 bp and 5014-5048 bp. The fragment corresponding to the 5' flank of the gene of interest is cloned into the former using the restriction sites *Sac*II, *Spe*I or *Afl*II. The 3' fragment is inserted into the second MCS using *Eco*RI, *Nco*I or *Avr*II restriction sites. pCC-1 includes the ampicillin resistance cassette (AmpR 5492-6352 bp) for selection in *E. coli* during cloning. For positive selection of parasites harbouring transfected plasmid, the vector contains a *human dehydrofolate reductase* (*hdhfr* 3881-4442 bp) resistance cassette under the control of the *P. falciparum* calmodulin promoter (PfCAM 5' 2890-3874 bp). At the 3' end the *P. falciparum* histidine rich protein II 3' UTR (PfhrpII 449-5015 bp) is present. For negative selection of parasites with plasmid integration into the genome, the vector contains the *S. cerevisiae cytosine deaminase* (*cyd*) gene as a bifunctional chimeric protein with *S. cerevisiae* uracil phosphoribosyl transferase (*cyd* 886-2007 bp) under the control of the *P. falciparum* heat shock protein 86 promoter with the *P. berghei dhfr* terminator (PbDT 3' 9-853 bp) at the 3' end.

2.5.3 Generation of chemically competent cells

A 5 ml overnight culture was incubated at 37ºC with shaking at 200 rpm. This was used to inoculate 500 ml (1/1000 dilution of overnight culture) of LB medium and the bacteria were grown at 37ºC with shaking at 200 rpm until mid–log phase (OD_{600} = 0.5 – 0.6). The culture was then cooled on ice for 15 min and the bacteria harvested by centrifugation at 4000 g for 10 min at 4ºC. The bacterial pellet was resuspended in 30 ml Tfb I and incubated on ice for 15 min. The cells were then spun down at 4000 g for 5 min, the supernatant was removed and the bacteria resuspended in 6 ml Tfb II. The now chemically competent cells were flash frozen in dry ice in 50 µl aliquots in 1 ml tubes and stored at -80ºC.

2.5.4 Transformation of chemically competent cells

A total of 1 µl of plasmid or 2-5 µl of ligation mixture was added to a 50 µl aliquot of chemically competent *E. coli*, which had been thawed on ice. The mixture was incubated for 20 min on ice, after which the cells were heat shocked in a 42ºC water bath for 45 s to induce insertion of DNA into the bacteria. The mixture was immediately cooled on ice for 2 min, which was followed by addition of 250 µl of SOC medium (section 2.3.4) pre-warmed to 37ºC. The mix was incubated at 37ºC with shaking at 200 rpm for 1 h to allow recovery of bacteria, after which aliquots of 50 µl and 200 µl were spread onto LB agar plates containing the appropriate antibiotic and any possible additive (IPTG/X-Gal) required for a cloning step (section 2.5.2). For chemically competent cells used for cloning and for protein expression see Table 2-5.

2.5.5 Colony cracking

This method, as described by Barnes (1977), is an easy way to screen a large number of colonies for the presence of plasmid DNA containing an insert. It is based on the differing migration of bacterial DNA compared with plasmid DNA with and without insert. On each occasion, 17 positive colonies (white) and 3 negative control colonies (blue) from a blue-white colony screening plate containing 2% (v/v) X-Gal were tested. Each colony was

transferred onto an agar plate with a grid reference (LB/Kan with X-Gal as in section 2.5.2.1, or LB/Amp with IPTG and X-Gal as in section 2.5.2.3), and then resuspended in 20 µl colony cracking lysis buffer (section 2.3.1). This was then incubated at 55°C for 30 min after which each suspension was mixed for 30 s to shear genomic DNA. A total of 4 µl DNA loading dye (New England Biolabs) was added and the mixtures were vortexed again for 30 s. The samples were centrifuged for 1 min at 16,000 g to pellet any cell debris. The DNA in the samples was then separated on a 0.8% agarose gel at 50 V for several hours and the the electrophoretic mobility of DNA isolated from potential positive colonies (white) and the DNA isolated from negative controls (blue colonies) was compared . Plasmid DNA bands were visible between the *E. coli* genomic DNA (20 – 30 kb) and the low molecular mass (~500 bp) RNA.

2.5.6 Purification of plasmid DNA from bacteria

Depending on the amount of plasmid DNA to be isolated, two different methods were used. Both methods involve lysis of bacteria using SDS, precipitation of bacterial proteins and DNA with alkaline solution and purification of DNA using silica columns.

For small-scale purification of plasmid DNA (generally for 10 - 30 µg DNA) the Qiaprep Spin Miniprep Kit (Qiagen) was used. A single colony from a bacterial transformation was used to inoculate 2-10 ml LB medium with the appropriate antibiotic. The cultures were grown overnight at 37ºC with shaking at 200 rpm. The bacteria were harvested by centrifugation at 16,000 g for 3 min and DNA purified as per manufacturer's instructions. Elution of DNA from the silica column was performed with 30 - 50 µl of distilled water (dH_2O).

The Qiaprep Spin Maxiprep Kit (Qiagen) was used for large-scale purification (~ 0.7 mg DNA). Overnight pre-cultures were prepared by inoculating 2 ml of LB/Amp with a single colony from a transformation reaction and grown at 37ºC with constant shaking at 200 rpm. This was used to inoculate 500 ml of TB/Amp, and the bacteria were grown overnight at 37ºC with shaking at 200 rpm. The bacteria were harvested by centrifugation at 4000 g for 15 min and the plasmid purified as described in the manufacturer's instructions. Precipitated DNA was redissolved in 200 µl of TE buffer (section 2.3.1).

2.5.7 Restriction endonuclease digests

Restriction endonuclease digests were performed using enzymes and their appropriate buffers from New England Biolabs. Digests were carried out in a 37ºC incubator to avoid loss of any sample due to evaporation. For restriction digests of plasmid, 2 - 6 µl of plasmid purification was digested for 1.5 - 2 h before electrophoresis. Digestion of genomic DNA (2 - 2.5 µg) was carried out overnight.

2.5.8 Agarose gel electrophoresis

For separation of digested DNA fragments, agarose gel electrophoresis was used. Agarose (0.8-1% w/v) was dissolved in 1x TAE buffer (section 2.3.1) and 0.5 µg/ml of Sybr Safe was added. Electrophoresis of DNA samples mixed with 6x DNA loading dye was carried out in 1x TAE buffer at 5-10 V/cm. A 1 kb DNA ladder was run next to the samples to determine the sizes of DNA fragments. DNA was visualised by taking a picture with a UV illuminator (GelDoc 2000, BioRad).

2.5.9 Gel extraction

For extraction of DNA from gels, low melting point agarose was used to separate the DNA fragments. A DarkReader transilluminator (Clare Chemical Companies) was used to visualise the DNA and a scalpel was used to cut the fragments out, trimming off as much of the surrounding gel as possible. The QIAQuick Gel Extraction Kit (Qiagen) was used to extract the DNA as per manufacturer's instructions. DNA was eluted in 30 - 50 µl of dH_2O.

2.5.10 Determining DNA concentration

DNA concentrations were determined spectrally at 260 nm using a NanoDrop 1000 operated with ND-1000 V.3.7.0 software (Thermo Scientific). The A_{260nm}:A_{280nm} and A_{230nm}:A_{260nm} ratios, respectively were recorded to determine purity of the DNA.

2.5.11 Southern blot

Southern blots were performed to determine the genotype of *P. falciparum* mutant and wild-type lines. A total of 2 – 2.5 µg of genomic DNA and 0.2 ng of the plasmid used for transfection were digested at 37ºC overnight with appropriate restriction enzymes in a 30 µl reaction volume. The digested DNA was mixed with 6x DNA loading buffer and separated on 0.8% agarose gel. The initial current was set at 100 V for 10 min to allow DNA to enter the gel, which was followed by separation of digested DNA at 40 – 50 V for 5 - 6 h. A 1 kb DNA ladder was included on both sides of the gel. Following separation, the gel, with a fluorescent ruler next to it, was photographed so that migration distances of DNA fragments could be determined in relation to the DNA standards on either side of the gel. The gel was soaked in 250 ml 0.25 M HCl for 10 min followed by 2 x 10 min wash with dH_2O. A further 2 x 15 min incubation in 250 ml of 0.4 M NaOH was carried out. Transfer of DNA to a Hybond-N+ nylon membrane (GE Healthcare) was performed overnight in a sandwich as shown in Figure 2-6.

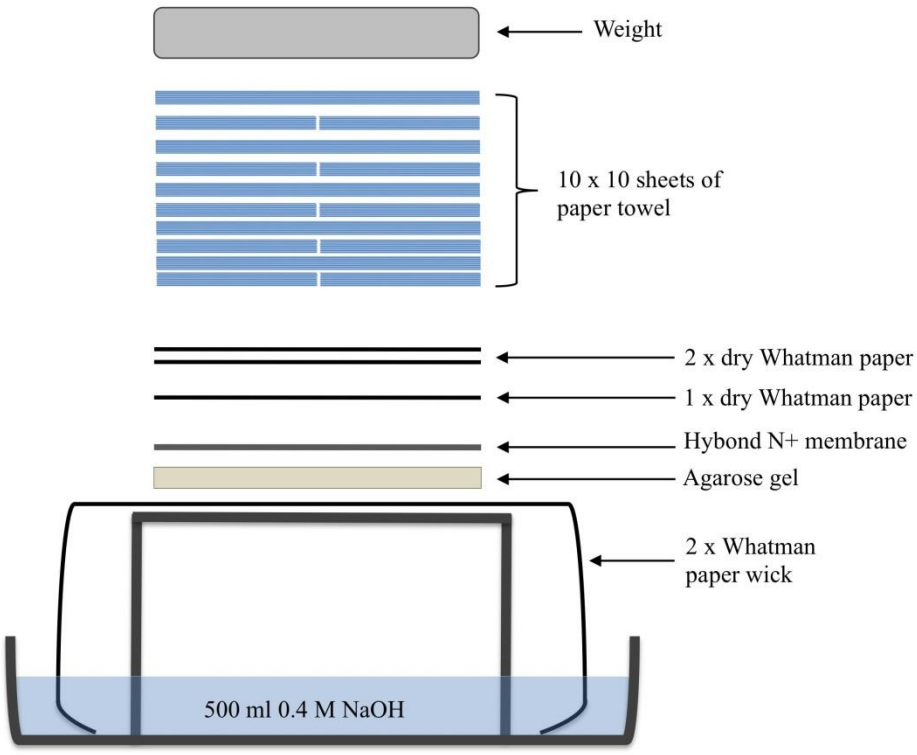


Figure 2-6. Southern blotting apparatus
Transfer of DNA to a Hybond-N+ nylon membrane was carried out using capillary blotting where the membrane is placed between the gel and absorbent paper towels to draw the 0.4 M NaOH through the gel using capillary action.

Following transfer of DNA, the wells were marked with a pencil and the membrane was washed in 2x SSC buffer (section 3.2.1) for 10 min and air-dried. The DNA was then fixed onto the membrane by UV cross-linking (254 nm) at 700 x 100 $\mu J/cm^2$).

The probes used to detect DNA fragments were either gene specific, or specific for the selectable markers of the pCC-1 plasmid (Fig. 2-5). The probes were generated by PCR or isolated by restriction digest from the pCC-1 transfection plasmid and labelled using the Gene Images AlkPhos Direct Labelling kit (GE Heathcare) as instructed by the manufacturer. The membrane was incubated overnight at 60ºC in hybridisation buffer (section 2.3.1) with the labelled probes. The membranes were then washed 3 x 15 min with 50 ml primary wash buffer (section 2.3.1) at 60ºC followed by 2 x 10 min with 100 ml secondary wash buffer (section 2.3.1) at room temperature. To visualise the probes, the membrane was incubated for 5 min with CDP Star detection solution (GE Healthcare) followed by exposure of the membrane to autoradiography film for 30 – 90 min (Amersham).

2.5.12 Stripping a Southern blot membrane

To strip a Southern blot membrane for re-probing, the membrane was incubated in 0.5% SDS for 1 h at 60ºC. The membrane was then washed twice in 2x SSC buffer (section 2.3.1) for 10 min and air-dried before incubation with the new probe as described above.

2.6 Biochemical methods

2.6.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

The Novex® NuPAGE® electrophoresis system (Life Technologies) was used for SDS-PAGE. For expression trials, gels were self-made and consisted of a 5% polyacrylamide stacking gel and 10% polyacrylamide resolving gel (for gel mix compositions see 2.3.2). To make the SDS-PAGE gels, 20 µl of ammonium persulphate (20% w/v) (Fisher Scientific) and 6 µl of N,N,N',N'-tetramethylethylenediamine (TEMED) (Sigma) were added to polymerise 6 ml of the resolving gel mix, which was poured into a Novex® 1 mm cassette (Life Technologies). A ~1 cm layer of butanol (Sigma) was added to ensure that the top of the resolving gel was level and to remove any air bubbles; the gel was left to polymerise for 30 min. The butanol was then washed off with dH_2O and 2 ml of stacking gel was mixed with 10 µl of ammonium persulphate and 3 µl of TEMED before it was poured on top of the separation gel. An appropriate comb was inserted immediately and the gel was allowed to set for at least 30 min before use. The gels were run in a Xcell Surelock™ Mini-cell apparatus with Tris/glycine resolving buffer (section 2.3.2) at a constant current of 20 mA until the protein had migrated through the stacking gel into the resolving gel. After this, the current was increased to 40 mA. During protein purification, pre-cast NuPAGE Bis/Tris 10% polyacrylamide gels were used with 1x NuPAGE® MES SDS Running Buffer (Life Technologies) as this combination allowed faster resolving of proteins between purification steps at a constant voltage of 200 V. Protein samples were prepared by denaturing for 10 min at 100ºC in 1x SDS loading dye (section 2.3.2). Following separation the gels were stained in GelCode Blue Safe Stain (Thermo Scientific) for 30 min to 1 h and washed in dH_2O for 2-3 h to destain.

2.6.2 Western blot analyses

A Trans-Blot® SD Semi-Dry Transfer Cell from Bio-Rad was used to transfer protein onto Protan nitrocellulose membrane (Whatman) according to manufacturer's instructions. Transfer was in a down-ward direction in a sandwich made up of five pieces of Whatman paper soaked in Towbin buffer (section 2.3.2) on the top followed by the SDS-PAGE gel, nitrocellulose membrane and another five Whatman paper pieces all soaked in Towbin

buffer. A fixed voltage of 20 V was applied, while the transfer time varied between 40 – 75 min, depending on the size of the protein and number of gels being transferred. The membranes were then stained with Ponceau S (Sigma) to confirm protein transfer and the lanes were marked with a pencil. Following removal of the Ponceau S with repeated washing with PBS containing 0.1% (v/v) Tween 20 (PBS/Tween20) the membrane was blocked on a shaker for either 1 h at room temperature or overnight at 4ºC in 5% (w/v) skimmed milk in PBS. After blocking, membranes were washed 3 x 10 min in PBS/Tween20. The membrane was then incubated for 1 h with shaking in primary antibody at the required dilution (Table 2-6) in 2% skimmed milk in PBS/Tween20. The membrane was again washed for 3 x 10 min in PBS/Tween20 and incubated for another 1 h in secondary antibody at the appropriate dilution (Table 2-6) in 2% skimmed milk in PBS/Tween20. Following the final 3 x 10 min wash in PBS/Tween20, the horseradish peroxidase (HRP) conjugated secondary antibodies were detected with Immobilon Western Chemiluminescent HRP Substrate (Millipore) and visualised by exposing blots to X-ray films (Kodak).

2.6.3 Stripping a western blot

To strip a western blot for analysis with another antibody, the membrane was first washed for 5 min in PBS and then sealed into a plastic bag with 10 ml of Restore western blot stripping buffer (Thermo Scientific) and incubated at 37ºC for 30 min. The membrane was washed for 5 min in PBS/Tween20 before resuming with blocking, washing and re-probing the membrane as described above (section 2.6.2).

2.6.4 Recombinant protein test expressions

To obtain the most favourable expression conditions small-scale expression trials were carried out. Single colonies were inoculated into 2 ml of LB/Amp medium and grown overnight at 37ºC with shaking at 200 rpm. A total of 50 µl of the overnight culture was added to 50 ml of LB/Amp medium and the cultures were grown at 37ºC with shaking at 220 rpm until mid–log phase (OD_{600nm} = 0.5 – 0.6). The cultures were then placed into incubators at various temperatures (15ºC, 18ºC, 25ºC, 30ºC and 37ºC) and expression was

induced with 0.5 – 5 mM of IPTG. For analysis of protein expression, 1 ml samples of the cultures were taken before induction (0 h) and after 1 – 4 h expression and after overnight expression. The samples were centrifuged at 16,000 g for 3 min to remove the medium and the pellets were stored at -20ºC awaiting further analysis.

2.6.5 BugBuster protein extraction

To extract the protein from the test expression experiment, the pellets were resuspended in 100 µl of BugBuster protein extraction reagent (Novagen), which contains a proprietary mixture of non-ionic detergents that disrupt bacterial cells without denaturing the protein. To degrade DNA and RNA, 0.5 µl benzonase (Novagen) was added and to prevent protein degradation protease inhibitors were added at the following final concentrations; 20 µM leupeptin, 2 µM pepstatin A, 1 mM phenylmethanesulphonyl fluoride (PMSF), 1 mM benzamidine, 2 µM 1,10-phenanthroline and 10 µM E-64. The samples were incubated at room temperature for 15 min with shaking at 1500 rpm (VXR basic Vibrax® with test tube insert, IKA) and centrifuged for 15 min at 16,000 g at 4ºC. The supernatant, which contained the soluble protein fraction, was transferred into a new 1 ml tube and the remaining pellet was resuspended in 100 µl of PBS with the above protease inhibitors. A 15 µl aliquot of the supernatant and 2 µl aliquot of the resuspended pellet were mixed with 1 x SDS loading dye, denatured at 100ºC for 10 min, separated by SDS-PAGE and analysed by western blotting.

2.6.6 Large-scale recombinant protein expression

Recombinant proteins were expressed in *E. coli*. Strains, temperature and time of expression and the amount of IPTG used for induction were optimised and are described in Table 2-7. Recombinant protein expression with a colony from a glycerol stock was poor, thus each time a freshly transformed bacterial colony was used to inoculate a 10 ml LB/Amp starter culture, which was incubated overnight at 37ºC with shaking at 220 rpm. For each expression batch, 8 x 500 ml of TB/Amp medium were inoculated with 1 ml of the overnight culture. The bacteria were grown at 37ºC with shaking at 220 rpm to an

OD_{600nm} of 0.6-0.8, after which the cultures were cooled to the optimum temperature for expression of each construct before induction with 0.5 - 1 mM IPTG. The cultures were

Table 2-7. Large-scale expression conditions
The host *E. coli* strain, temperature, time and concentration of IPTG used for induction of each construct. Also shown are the buffers in which the harvested bacterial pellets were resuspended (Table 2-1 for buffer compositions).

Construct	Host *E. coli*	Temp ºC	Time	[IPTG] (mM)	Resuspension buffer
His-r*Pf*E2m	Rosetta™ 2(DE3)pLysS	18	overnight	0.5	POROS binding buffer
His-r*Pf*E2bc	BLR(DE3)	25	overnight	1	HisTrap binding buffer
GST-r*Pf*E2c	BLR(DE3)	18	overnight	1	GST binding buffer
His-r*Pf*aE3	NovaBlue(DE3)	30	overnight	1	Qiagen lysis buffer

then incubated overnight at the appropriate temperature with constant shaking at 220 rpm. Bacteria were harvested by centrifugation at 3000 g for 15 min at 4ºC. The pellets were resuspended in the appropriate buffer required for the purification method to be used (Table. 2-7). Protease inhibitors (20 μM leupeptin, 2 μM pepstatin A, 1 mM PMSF, 1 mM benzamidine, 2 μM 1,10-phenanthroline and E-64 10 μM) were added to the resuspended pellets and stored at -20ºC until purification. For His-r*Pf*aE3 overexpression, flavin adenine dinucleotide (FAD) was added to a final concentration of 100 μM.

2.6.7 Protein extraction for purification

Bacterial pellets were defrosted at room temperature and 50 μg/ml lysozyme (Sigma) was added to lyse bacterial cell walls, 1 μg/ml of DNase (Roche) was added to degrade DNA and RNA, and to prevent protein degradation, protease inhibitors (20 μM leupeptin, 2 μM pepstatin A, 1 mM PMSF, 1 mM benzamidine, 2 μM 1,10-phenanthroline and 10 μM E-64) were also added. The mixture was incubated on ice for 30 min after which the OneShot (PSI) programme set to 15 kpsi on the OneShot Cell Disrupter was used to disrupt the bacteria. The lysate was centrifuged at 48,000 *g* for 1 h, and filtered through a 45 μm Millex®-HV PVDF syringe filter (Millipore) before use in various purification columns.

2.6.8 Recombinant protein purification

Recombinant protein was purified on an ÄKTApurifier controlled by UNICORN 5.1 software (both GE Healthcare). Programmes were set on the UNICORN software to load, wash and elute protein from the column. All buffers used are listed in Table 2-1. The buffers, dH_2O and 20% ethanol were passed though a 0.22 μm filter and degassed for at least 30 min before use.

2.6.8.1 Nickel HisTrapFF

HisTrap FF columns are packed with Ni Sepharose 6 Fast Flow medium, which contains 90 μm highly cross-linked agarose beads with a chelating group to bind metal ions. The column was used at a flow rate of 1 ml/min. On its first use, to prepare a 1 ml HisTrap FF (GE Healthcare) column charged with Ni^{2+} for purification, the column was washed with 10 column volumes (CV) of dH_2O and equilibrated with 5 CV of HisTrap binding buffer. When the column was reused, binding capacity was maintained by washing the HisTrap FF with 10 CV of dH_2O, stripping the metal ions with 10 CV of Stripping buffer followed by a wash with 10 CV of HisTrap binding buffer followed by 10 CV of dH_2O. The column was then recharged with 0.5 ml 0.1 M $NiSO_4$, washed with 5 CV of dH_2O and equilibrated with 5 CV of HisTrap binding buffer before use. A total of 15 ml of filtered protein lysate was loaded onto the column and the flow-through collected into a separate tube. HisTrap buffer A (no imidazole) and buffer B1 and B2 (1 and 2 M imidazole, respectively) were mixed by the ÄKTApurifier to obtain the required imidazole concentration for each wash and elution step. An initial 2 CV wash with 10 mM imidazole was followed by a second wash for 5 – 10 CV with 40 mM imidazole. Elution was either with a step-wise increase in imidazole concentration (100 mM, 400 mM and 1 M) as for His-r*Pf*E2m or by a linear imidazole concentration gradient (40 mM - 2 M) as for His-r*Pf*E2bc. The wash and elution steps were collected in 1 ml fractions and were analysed by SDS-PAGE and western blot along with the flow-through. All the samples were stored at 4ºC. After purification the column was washed with 10 CV of dH_2O and 10 CV of 20% ethanol in which it was stored in at room temperature.

2.6.8.2 Zinc POROS® PEEK™ MC

A POROS® PEEK™ 1.7 ml column was manually packed in-house with POROS® MC 20 µm Self Pack® Media using the ÄKTApurifier. The column was stripped of metal ions prior to each purification with 10 CV of Stripping buffer, washed with 10 CV of dH_2O and re-charged with 30 CV of 0.25 M $ZnSO_4$. This was followed by a 10 CV wash with dH_2O followed by 10 CV of 0.25 M NaCl to remove any ionically bound metals. To improve selectivity and binding, the column was saturated with imidazole (5 CV of POROS buffer B with 1 M imidazole). This displaces water molecules on the metal interaction sites with the elution agent, resulting in a reduction in binding strength, which improves selectivity. The final step was to equilibrate the column with 10 CV of POROS binding buffer (10 mM imidazole). The column preparation steps were carried out at a flow rate of 10 ml/min, while the following purification steps were performed at 5 ml/ml. Protein lysate was filtered and 15 ml loaded onto the column and the flow-through collected for analysis. POROS Buffer A (no imidazole) and Buffer B (1 M imidazole) were mixed by the ÄKTApurifier to obtain the required imidazole concentration for each wash and elution step. A preliminary 5 CV wash with 10 mM imidazole was followed by elution of protein with a step-wise increase in imidazole concentration (100 mM, 500 mM and 1 M). Elution fractions (1.7 ml) were collected, and together with the flow-through analysed by SDS-PAGE and western blot. Following purification the column was washed with 10 CV of dH_2O followed by 10 CV of 20% ethanol in which it was stored in at room temperature.

2.6.8.3 Ni-NTA batch purification

His-r*Pf*aE3 was purified using a batch purification method with Ni^{2+}- nitrilotriacetic acid (Ni-NTA) agarose (Qiagen). Filtered bacterial lysate was incubated with 1 ml of 50% Ni-NTA resin per litre of bacterial culture for 1 h at 4ºC with rotation on a blood wheel. The mix was then poured into an Econo-Pac® chromatography column (Bio-Rad) and the flow-through was collected by gravity flow. The Ni-NTA resin was washed twice with 4 ml of Qiagen wash buffer 1 (20 mM imidazole) and once with 4 ml of Qiagen wash buffer 2 (50 mM imidazole). His-r*Pf*aE3 was eluted from the Ni-NTA beads with 1 ml aliquots of Qiagen elution buffer (250 mM imidazole) and collected by gravity flow. The flow-through, wash and elution fractions were stored at 4ºC and analysed by SDS-PAGE and western blot. The Ni-NTA agarose was not reused for subsequent purifications.

2.6.8.4 GSTrap FF

For purification of GST-r*Pf*E2c, GE Healthcare GSTrap FF 1 ml columns were used. The column is made of polypropylene with glutathione bound via a 10-carbon linker arm. Glutathione is the substrate for GST, thus it will bind GST-tagged proteins. Elution is achieved by displacing the GST-tagged protein with buffer containing reduced glutathione. The column was equilibrated with 5 CV of GSTrap binding buffer at a flow rate of 1 ml/min. A total of 10 ml of filtered protein lysate was loaded onto the column at a flow rate of 0.2 ml/min to maximise the interaction time between the GST-tag and the glutathione-coated beads, as the binding kinetics are slow. The flow-through was collected into a separate tube. The column was washed with 10 CV of GSTrap binding buffer followed by elution of GST-r*Pf*E2c with GSTrap elution buffer (10 mM reduced glutathione). The wash and elution steps were collected as 1 ml fractions. The flow-through, wash and elution fractions were analysed by SDS-PAGE and western blotting.

2.6.8.5 GST-tag removal

The GST-tag on pGEX-6P-1 was cleaved using PreScission Protease (GE Heathcare). Cleavage was attempted both on eluted GST-r*Pf*E2c and while the recombinant protein was still bound to the GSTrap column. For the former method, the reduced glutathione in the eluted GST-r*Pf*E2c was removed by buffer exchange into PreScission cleavage buffer using a PD-10 column (GE Healthcare). The protein was incubated overnight at 4ºC with 30 µl (60 U) of PreScission Protease and loaded back onto an equilibrated GSTrap column to separate the GST-tag free protein. The GST molecule and PreScission protease should bind onto the column and the recombinant protein elute in the flow-through. For on-column cleavage of the tag, GST-r*Pf*E2c was not eluted from the GSTrap column. Instead following loading of protein, the column was washed with 10 CV of PreScission cleavage buffer. The PreScission protease (30 µl (60 U) in 970 µl PreScission cleavage buffer) was loaded onto the column using a syringe and the column incubated overnight at 4ºC. A syringe was used to elute the cleaved recombinant protein in 3 x 1 ml fractions of PreScission cleavage buffer. All fractions were analysed using SDS-PAGE and western blotting.

2.6.8.6 Gel filtration

Gel filtration was carried out with HiLoad™ 16/60 Superdex 200 and HiPrep 26/60 Sephacryl S-400 HR gel filtration columns at 4ºC. The columns were washed with 2 CV dH_2O and equilibrated with 2 CV of the appropriate buffer (Table 2-1). A maximum volume of 5 ml and 10 ml of protein was applied to the Superdex 200 and Sephacryl S-400 columns, respectively. Elution was carried out with 1.5 CV of the appropriate buffer at a flow rate of 0.75 ml/min for the Superdex 200 column and 1.3 ml/min for the Sephacryl S-400 column. Elutions were collected in 1 ml fractions. Following purification, the columns were washed with 2 CV dH_2O and 3 CV 20% ethanol in which they were stored in at 4ºC. Standards of known molecular mass were separated on both columns to generate a standard curve from which the molecular masses of the recombinant proteins could be estimated. The standards used were dextran blue (2 MDa, 0.6 mg/ml), thyroglobulin (669 kDa, 2.5 mg/ml), apoferritin (440 kDa, 2.5 mg/ml), alcohol dehydrogenase (150 kDa, 2.5 mg/ml) and bovine serum albumin (66 kDa, 2.5 mg/ml).

2.6.9 Determination of protein concentration

Protein concentrations were determined using the Bradford assay with bovine serum albumin as a standard (Bradford, 1976) using the 5 x Bio-Rad assay reagent. One part assay reagent was added to four parts of standard protein at concentrations ranging from 0.01 mg/ml to 1 mg/ml and the absorbance was measured at 595 nm. The unknown concentration of recombinant protein was calculated from a standard curve generated with the known concentrations of BSA. Concentration of the recombinant proteins was also determined spectrally at 280 nm with a NanoDrop 1000 operated with ND-1000 V.3.7.0 software (Thermo Scientific) and using the appropriate extinction coefficient (His-r*Pf*E2m = 22 015 M^{-1} cm^{-1}, His-r*Pf*E2bc = 46 488 M^{-1} cm^{-1}, His-r*Pf*aE3 = 44 950 M^{-1} cm^{-1} calculated from the amino acid composition using PROTPARAM (Gasteiger *et al*., 2005, http://web.expasy.org/protparam/).

2.6.10 Concentrating recombinant protein

Concentration of recombinant protein was carried out either with 2 ml Vivaspin™ sample concentrators (GE Healthcare) with a 30 kDa molecular mass cut-off or with Aquacide II (Millipore). The concentrators were spun at speeds appropriate to the centrifuge rotor used (4000 g for the swing out rotor) until the desired concentration of protein was obtained. Aquacide II was used as an alternative method for concentrating His-r*Pf*E2m, which bound onto the membrane of Vivaspin™ concentrators. Aquacide is a sodium salt of carboxymethylcellulose, which absorbs water and can thus be used to remove water from protein samples in dialysis tubing.

2.6.11 Buffer exchange using a PD-10 column

PD-10 desalting columns (GE Healthcare) were used for buffer exchange of protein samples. The column was equilibrated with 25 ml of the desired final buffer after which a maximum of 2.5 ml of the sample was added to the column. The flow-though was discarded and the protein was eluted with 3.5 ml of the required buffer and collected in 1 ml fractions. The protein was usually found in the second and third fraction.

2.6.12 Dialysis

Following concentration with Aquacide II the protein sample was dialysed into the required buffer. Sample was placed into SnakeSkin® 10 K MWCO 16 mm dry dialysis tubing and dialysed in 1000x volume of the sample overnight at 4ºC.

2.6.13 E2 activity assay

The activity of His-r*Pf*E2m and His-r*Pf*E2bc was assayed as described by Yang *et al*. (1997) with modifications. The assay couples the production of acetyl-CoA with the reverse reaction catalysed by E2. The reaction was followed by measuring the increase in absorbance at 240 nm due to production of 8-acetyl-dihydrolipoamide, in a UV-2501 spectrophotometer (Shimadzu). The reaction was carried out at 25ºC in 100 mM Tris-HCl,

1 mM EDTA, pH 7.4 with 1 mM acetyl-phosphate and 2 units of phosphotransacetylase. The amounts of coenzyme-A (CoA), dihydrolipoamide and His-r*Pf*E2m or His-r*Pf*E2bc were varied depending on the parameters analysed. To obtain a steady starting absorbance reading, His-r*Pf*E2m and His-r*Pf*E2bc were added last into the reaction mix, because all the other components absorb light at 240 nm.

For measuring the activity of various concentrations of His-r*Pf*E2m, 0.3 – 2 μg of protein was used. For all other assays 1 – 2 μg of His-r*Pf*E2m or His-r*Pf*E2bc was added to the reaction. To determine the K_m and v_{max} for CoA, dihydrolipoamide was kept constant at 1 mM, while the concentration of CoA was varied between 1 μM – 50 μM. To measure the optimal amount of dihydrolipoamide required for the assay, CoA was kept constant at 10 μM and dihydrolipoamide was varied between 0.1 – 5.0 mM.

2.6.14 E3 activity assays

Both the forward and reverse reaction of E3 was assayed to determine the kinetic parameters of His-r*Pf*aE3 as described by McMillan *et al*. (2005). The production or consumption of NADH was measured at 340 nm in UV-2501 spectrophotometer (Shimadzu).

2.6.14.1 Forward reaction

The substrates for the forward reaction are dihydrolipoamide and NAD^+. The reaction was performed in 50 mM potassium phosphate, 1 mM EDTA, pH 8 with 1 μg of His-r*Pf*aE3 at 25ºC. Dihydrolipoamide (Sigma) was freshly prepared as a 60 mM stock with 100% ethanol and NAD^+ was prepared with reaction buffer as a 10 mM stock solution. The K_m and v_{max} for dihydrolipoamide were determined by keeping NAD^+ constant at 2 mM and varying the concentration of dihydrolipoamide between 50 μM and 2 mM. Likewise for the kinetic parameters for NAD^+, dihydrolipoamide was kept constant at 2 mM and NAD^+ varied between 62.5 μM and 2 mM.

2.6.14.2 Reverse reaction

The substrates for the reverse reaction are lipoamide and NADH. The reaction was performed in 50 mM potassium phosphate, 1 mM EDTA, pH 7, at 25ºC with 1 μg of His-r*Pf*aE3. Lipoamide (Sigma) was freshly prepared with 100% ethanol as a 60 mM stock, which required heating at 37ºC and repeated vortexing to dissolve. NADH was prepared in reaction buffer as a 10 mM stock solution. To determine the K_m and v_{max} for lipoamide the concentration of NADH was kept constant at 200 μM and lipoamide was varied between 50 μM and 4 mM. For the kinetic parameters for NADH, lipoamide was kept constant at 2 mM and NADH was varied between 10 μM and 200 μM.

2.7 Biophysical methods

2.7.1 Protein partial specific volume and buffer density and viscosity calculations

SEDNTERP (Laue *et al.*, 1992) (http://bitcwiki.sr.unh.edu/index.php/Main_Page) was used to calculate the partial specific volume ($\bar{v}$) of each recombinant protein from their amino acid compositions (Table 2-8). Also the density (ρ) and viscosity (η) of each buffer used for AUC experiments was calculated (Table 2-9).

Table 2-8. Calculated recombinant protein partial specific volumes at 4ºC and 20ºC.

	$\bar{v}$ (ml/g)	
Protein	**4ºC**	**20ºC**
His-r*Pf*E2m	0.7338	0.7406
His-r*Pf*E2bc	0.7356	0.7424
His-r*Pf*aE3	0.7317	0.7385

Table 2-9. Buffer parameters used for analysis of AUC data.
The composition of buffers PEBS150, PEBS250 and PEBS150+glycerol+DTT can be seen in Table 2-1. The composition of the other buffers is the same as PEBS150 and PEBS250 – the number at the end denotes the concentration of NaCl (mM) in the buffer.

Buffer	Density (ρ, g/ml)	Viscosity (η, cPoise)
PEBS0	1.00774	1.5985
PEBS10	1.00815	1.6022
PEBS50	1.00981	1.6073
PEBS100	1.01188	1.6141
PEBS150	1.01394	1.5670
PEBS250	1.01804	1.6354
PEBS500	1.02821	1.6820
PEBS1000	1.04815	1.7496
PEBS150+glycerol+DTT	1.01414	1.5670

2.7.2 Sedimentation velocity (SV)

Sedimentation velocity experiments were carried out using a Beckman Coulter Optima XL-I analytical ultracentrifuge (Palo Alto, USA) with an An-50 Ti 8-hole or An-60 Ti 4-hole rotor. Samples and reference buffers (360 μl) were loaded into double sector 12 mm path length centrepieces made of charcoal-filled Epon. The rotor with the cells was placed in the centrifuge and cooled under vacuum for several hours until the experimental temperature, 4ºC, was reached. Samples were spun at 3,000 or 49,000 rpm depending on the anticipated size of the protein in the sample. Both absorbance and interference optics were used to record data over a radial range of 6.0 – 7.25 cm. For absorbance data a radial step size of 0.002 cm was used and for interference optics the laser delay and fringe contrast and brightness at the experimental speed was adjusted before the run to obtain high quality fringes. A cell was scanned every 4 or 8 min depending on whether the 4- or 8-hole rotor was used. A total of 200 scans were taken. SEDFIT (Schuck, 2000; Schuck *et al.*, 2002) (http://www.analyticalultracentrifugation.com/default.htm) was used to determine the number of species in the sample and to obtain the sedimentation coefficients using the continuous distribution c(s) Lamm equation model (section 1.6.1.3). Firstly, the meniscus, cell bottom and range of data to fit were determined. Following this the parameters for the analysis were defined. The resolution of the curve was set to 200 and the range of sedimentation coefficients to be shown were set at 0.1-10 S after confirming no larger species were observed. The frictional ratio was not fixed. The appropriate partial specific volume, buffer density and viscosity were defined as shown in Table 2-8 and 2-9. The other parameters were kept at default values. The obtained sedimentation coefficients

were then standardised to $s_{20,w}$ values as described in section 1.6.1.1 and plotted against protein concentration to obtain the sedimentation coefficient at infinite dilution ($s^{0}_{20,w}$) from the y-intercept. Data from absorbance optics were of higher quality throughout, thus only absorbance data were used for the final analyses presented in this thesis.

2.7.3 Sedimentation equilibrium (SE)

Sedimentation equilibrium experiments were carried out using a Beckman Coulter Optima XL-I analytical ultracentrifuge (Palo Alto, USA) with an An-50 Ti 8-hole or An-60 Ti 4-hole rotor. A total of 80 µl of sample and reference buffer were loaded into double sector 12 mm path length, charcoal-filled, Epon centrepieces. The samples were centrifuged at 4ºC at a range of speeds depending on the protein size anticipated from sedimentation coefficients (Table 2-10).

Table 2-10. Speeds used for SE experiments

Protein	Speed (rpm)
His-r*Pf*E2m	12,000
His-r*Pf*E2bc	35,000
His-r*Pf*aE3	17,000

Ten scans, using absorbance and interference optics, were recorded over a radial range of 6.8 – 7.25 cm and were taken 3 h apart following a 3 h initial delay. WinMATCH (www.biotech.uconn/edu/auf/) was used to confirm that equilibrium had been reached in the sample, which was when no significant change in the root mean square deviation (rmsd $< 0.1 \times 10^{-2}$) was measured. SE data were analysed using SEDPHAT (Schuck, 2003; Vistica *et al.*, 2004) (http://www.analyticalultracentrifugation.com/sedphat/sedphat.htm). To obtain the apparent whole-cell weight average molecular mass, $M_{w,app}$, single fit analysis with one species was carried out for each sample. The model also allows fitting of up to four species, which was used to determine the molecular masses of all the species present in the samples. The user defined experimental parameters were the partial specific volume, buffer density and viscosity and shown in Table 2-8 and 2-9, and the experimental temperature. The meniscus, cell bottom and range of data to fit were determined before analysis. $M_{w,app}$ was plotted against protein concentration and from the y-intercept the

whole-cell average molecular mass at infinite dilution, M_w^0, was determined. The species analysis with mass conservation constraints model was used to further analyse His-r*Pf*E2m. This model allows defining the concentration and molecular mass of a specific species in a sample to aid determination of the correct molecular mass of another species.

2.7.4 Small angle X-ray scattering (SAXS)

Small angle X-ray scattering experiments were carried out at the EMBL/DESY), Hamburg, beamline X33. The beamline has been recently decommissioned, however, X33 was a bending magnet beamline in the DORIS III storage ring, which operates at beam currents of 100 - 140 mA with an X-ray wavelength of 0.15 nm and an electron energy of 4.5 GeV. The samples were measured at various concentrations in a vacuum cell with mica windows at 10ºC. The beamline had an automated sample changer allowing 8 samples and up to 3 buffers to be measured without interruption. The sample to detector distance was 2.7 m at which the resolution of the SAXS data is $\mathbf{s} = 0.06 – 6\ nm^{-1}$, (momentum transfer $\mathbf{s} = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength). Data were collected with a 2D photon counting Pilatus 1M-W pixel X-ray detector, which was calibrated with a standard sample of ~ 5 mg/ml of BSA prior to measuring experimental samples. The exposure time of the sample to the X-rays was 15 s and 8 frames were taken for each sample, which took approximately 2 minutes per sample. The raw data were fed into the automated SAXS data processing pipline (Petoukhov *et al.*, 2007) for initial averaging of sample data frames and buffer subtractions. The ATSAS 2.5 programme suite (Petoukhov *et al.*, 2012) (http://www.embl-hamburg.de/biosaxs/software.html) was used for all processing and analysis of SAXS data. PRIMUS (Konarev *et al.*, 2003) was used to assess the data for aggregation and to obtain the Guinier range and Kratky plot (section 1.6.2.3). The processed scattering curve was then used for determination of the *p(r)* distance distribution function and the maximum dimensions (D_{max}) using GNOM (Svergun, 1992). R_g was determined both from the Guinier approximation and from GNOM, while the molecular wmass of the samples were calculated from the Porod volume determined from the GNOM processed file using DATPOROD (Petoukhov *et al.*, 2007).

2.8 Bioinformatics and computational methods

2.8.1 Amino acid sequence alignments

Amino acid sequences were aligned using the multiple sequence alignment tool ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Larkin *et al.*, 2007; Goujon *et al.*, 2010). For all alignments default parameters were used.

2.8.2 Homology modelling

Homology modelling was carried out using the I-TASSER server for protein structure and function predictions (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Zhang, 2008; Roy *et al.*, 2010). The amino acid sequences of proteins were submitted in FASTA format and default parameters were chosen. I-TASSER generates several models in PDB (Brookhaven Protein Data Bank) format, which are given a 'C-score' based on how good the prediction is. The C-score ranges from -5 to 2 and the closer the value was to 2, the better the prediction was and the model was chosen for use in subsequent applications.

2.8.3 *Ab initio* modelling of SAXS data

Ab initio modelling of proteins was carried out using DAMMIF (Franke and Svergun, 2009). The programme uses spheres, or dummy atoms, to model the scattering envelope using a simulated annealing method (Svergun, 1999) (section 1.6.2.4). DAMMIF was used on the ATSAS online server (http://www.embl-hamburg.de/biosaxs/atsas-online/dammif.php) (Petoukhov *et al.*, 2007) with the input file being the p(r) distance distribution file generated by GNOM (section 2.7.4). The search volume was in nm^3 units, the P2 and P3 symmetry parameter was given for His-r*Pf*aE3 and His-r*Pf*E2bc, respectively. The programme was asked to run 20 times to obtain 20 models. Usually, generated models are clustered and averaged to obtain the final representative *ab initio* model. The online version of DAMMIF automatically does this; it groups the models it has generated into clusters based on their similarity using DAMCLUST (Kelley *et al.*, 1996; Petoukhov *et al.*, 2012). The models within these clusters are subsequently analysed by

DAMAVER (Volkov and Svergun, 2003) to obtain the representative model i.e. the model with the lowest average normalised spatial discrepancy (NSD) value (section 2.8.5).

Ab initio modelling of flexible proteins cannot be done with DAMMIF, which is appropriate for macromolecules with more rigid structures. To model His-r*Pf*E2bc better the Ensemble Optimisation Method (EOM) was used (Bernado *et al.*, 2007; Petoukhov *et al.*, 2012). The programme can be used to model intrinsically unfolded proteins or multi-domain proteins with flexible linkers. The method is based on the assumption that the flexible protein can be in an average number of conformations, which represent the experimental data (section 1.6.2.4). The programme first generates a pool containing a large number of possible conformations (the default is 10,000), which cover the experimental configurational space. From this, a genetic algorithm is used to divide the appropriate models into subsets to generate an ensemble of models, which fit the experimental data. The flexibility of His-r*Pf*E2bc arises from the 31 amino acid linker between the E1/E3 subunit binding domain (SBD) and the catalytic domain (CD). The files given to EOM in order of input were (1) a .seq file with the amino acid sequence of the whole protein, (2) the PDB file for the BD in monomeric form, (3) the PDB file for the CD in trimeric form, and (4) the experimental scattering curve .dat file. EOM was run with default parameters, apart from specifying P3 symmetry and fixing the position of the CD trimer to only allow movement of the linker and SBD to generate the models. A chi squared value (χ^2) determined how well the ensemble fit the experimental data.

2.8.4 Superimposition of *ab initio* models and homology models

Superimposition of the high-resolution homology models with the low-resolution scattering envelopes generated by *ab initio* modelling was done using the programme SUPCOMB (Kozin and Svergun, 2001). The programme can also superimpose low-resolution models with each other. SUPCOMB matches models by treating them as ensembles of points in three-dimensional space. Superimposition is done by minimising the dissimilarity measure, called the normalised spatial discrepancy (NSD), of the models. The alignment is started from the inertia-axis followed by minimisation of the NSD, where the minimal distance value between one point in the first model and all points in the second model is determined. This is then repeated *vice versa* with points in the second model.

These values are then added and normalised against the average distances between the neighbouring points for the two models to obtain a final NSD value, which is an indicator of how similar the models are (Koch *et al.*, 2003). Generally, NSD values of around one indicate that the two models are similar. Default values for programme parameters were used when running the programme.

2.8.5 Hydrodynamic modelling

US-SOMO (Rai *et al.*, 2005; Brookes *et al.*, 2010a; Brookes *et al.*, 2010b) was used to calculate the hydrodynamic parameters of the various homology and *ab initio* models obtained. For the calculations, each amino acid of the homology models is represented by two beads; one for the main chain component and the other for the side-chain. The volumes of the beads are determined by the atoms in the residue and by any water molecules bound onto them to account for protein hydration in solution (Brookes *et al.*, 2010b). The position of the beads corresponds to the residues they represent and any overlap is removed because the hydrodynamic interaction tensor is not valid for overlapping beads of differing radii. From the PBD files of the models, US-SOMO calculates, among other parameters, the estimated sedimentation and frictional coefficients and R_g, which can then be compared with the experimental data.

The hydrodynamic parameters of the EOM models were computed with the same bead modelling procedure as above. The DAMMIF models, however, were first processed using the AtoB ‘grid’ method (Byron, 1997) in order to remove the slight overlap between the constituent monomers at the dimer interface. In this method a cuboid with the dimensions of a molecule is created and divided into cubes, which in this case had dimensions of 5 Å. To create the hydrodynamic bead model a sphere with a radius proportional to the mass contained in the cube is then placed at the cube's centre of gravity (Byron, 2008).

Discrete molecular dynamics (DMD) (Dokholyan *et al.*, 1998; Ding and Dokholyan, 2006) simulations were carried out to model flexible sequences in homology models. The programme generates several models with possible conformations of the flexible region. The static regions were defined and default parameters were used apart from changing the temperature to 1 kcal/mol/kB, extending the run time to 50000 and the pdb output timestep was set to 1000 to create 50 pdb model files. The simulation was carried out on the Alamo

cluster via the University of Texas Health Sciences Centre at San Antonio. The models were then submitted into US-SOMO to compute the hydrodynamic parameters.

2.9 *P. falciparum* culture

2.9.1 Culture of *P. falciparum* erythrocytic stages

P. falciparum was cultured as described by Trager and Jensen, (1976). All work was carried out in a containment level 3 culture suite. All solutions and media were filter sterilised (0.2 µm filters) prior to use. Parasites were grown at 37ºC in sterile tissue culture flasks in a gas atmosphere of 5% CO_2, 1% O_2 and 94% N_2 or in a 5% CO_2 incubator. Erythrocytes used for culture were separated from buffy coat by centrifugation at 600 g for 15 min followed by at least 3 x 15 min washes (at 600 g) with blood wash medium (section 2.3.5) until the supernatant was clear. The washed erythrocytes were stored at 4ºC and used for a maximum of 4 weeks. Parasites were grown in 5 or 10 ml of complete RPMI 1640 medium (section 2.3.5) at 5% haematocrit. For general maintenance, medium was changed every 24 h, without disturbing the erythrocyte layer. Parasitemia was monitored by examining Giemsa stained blood smears and was kept around 1 – 4% by regularly diluting the culture with fresh erythrocytes.

2.9.2 Giemsa staining of blood smears

Blood smears were made by spreading a small drop (5 to 10 µl) of erythrocytes onto a glass slide. The blood was allowed to dry and then fixed with 100% methanol for at least 20 s. The slides were then stained for 5 -10 min in a 1:25 dilution of Giemsa stain (VWR) in H_2O and excess stain washed off with water. The slides were air-dried and examined using a light microscope with a 100x oil immersion objective. To determine the parasitemia, the number of infected erythrocytes in 1000 red blood cells was determined.

2.9.3 Thawing *P. falciparum*

Stabilates were taken from liquid nitrogen and thawed in the incubator at 37ºC. One tenth of the stabilate volume of thawing solution 1 (section 2.3.5) was slowly added while carefully mixing and the cell suspension was incubated at room temperature for 5 min. Ten times the original stabilate volume of thawing solution 2 (section 2.3.5) was added slowly while gently mixing and the suspension was incubated for another 5 min at room temperature. Finally, 10 times the original stabilate volume of thawing solution 3 (section 2.3.5) was slowly added, mixed and incubated at room temperature for 5 min. The cell suspensions were centrifuged at 800 g for 5 min. The supernatant was carefully discarded and the parasites were transferred into a flask with 9.3 ml culture medium and 200 µl fresh erythrocytes.

2.9.4 Freezing *P. falciparum* stabilates

To prepare three parasite stabilates, a 10 ml culture with at least 2% ring stage parasites was centrifuged at 800 g for 5 min at 4ºC and the medium was discarded. The erythrocytes were resuspended in 1x pellet volume of complete RPMI 1640 medium followed by 2 x pellet volume of cold freezing solution (PBS with 30% (v/v) glycerol). The mixture was aliquoted (600 µl) into cryotubes and placed on ice for 15 min before transfer into liquid nitrogen.

2.9.5 Synchronisation of *P. falciparum* cultures

Sorbitol was used to synchronise ring-stage parasites as described by Lambros and Vanderberg, (1979). Cultures were centrifuged at room temperature at 800 g for 5 min. The medium was removed and the erythrocytes mixed with 5x pellet volumes of sorbitol (5% (w/v) in 10 mM potassium phosphate pH 7.2) and incubated for 5 min at 37ºC. The mixture was centrifuged again as above and the supernatant removed before returning the parasites into 10 ml of fresh pre-warmed (to 37ºC) complete RPMI 1640 medium.

2.9.6 Isolation of *P. falciparum* from erythrocytes

To isolate DNA and protein from *P. falciparum*, large cultures (50 ml) were allowed to reach a parasitemia of ~ 10% and the parasites were freed from contaminating erythrocytes by saponin lysis. Approximately 5 – 10 ml of medium was aspirated from the flask and the erythrocytes were resuspended in the medium. Saponin (2% (w/v) in PBS) was added at a tenth of the remaining culture volume. For complete lysis of erythrocytes the cells were incubated on ice for 10 min and the parasites harvested by centrifugation at 1800 g for 10 m in at 4ºC. The supernatant was removed and the parasite pellet washed twice by centrifugation as above in 10 ml cold PBS. The pellet was then resuspended in 1 ml cold PBS and transferred into a 1.5 ml tube and centrifuged at 2,300 g for 5 min. The supernatant was removed and the pellet was washed again in 1.5 ml cold PBS until the supernatant was clear. The parasite pellet was then used immediately for DNA extraction (section 2.9.7).

2.9.7 DNA extraction from *P. falciparum*

A phenol extraction method was used to isolate genomic DNA from *P. falciparum*. The isolated parasite pellet (usually 100 - 200 µl), was resuspended in 50 mM Tris HCl, pH 9.0, 0.2 M NaCl, 0.1 M EDTA and 1 mg/ml proteinase K by repeated pipetting. Finally, 1% (w/v) SDS was carefully mixed in to prevent excessive foaming and the mixture was incubated overnight at 50ºC with rotation. The following morning, 1 ml of phenol/chloroform was added to the mixture and incubated at room temperature for 1 h with rotation. During this time, proteins were denatured due to the phenol/chloroform and were extracted into the organic solvent, while DNA stayed in the aqueous buffer. The aqueous layer was separated from the phenol/chloroform by centrifugation for 10 min at 3000 g and the upper layer was carefully collected into a fresh tube with a Pasteur pipette. The phenol/chloroform extraction was repeated and the upper layer collected again. DNA was precipitated by addition of 1/30 sample volume of 3 M sodium acetate, pH 5 and 0.6x sample volume of isopropanol. The mixture was inverted several times and the precipitated DNA spooled out and dissolved in 200 µl of TE buffer (section 2.3.1).

2.9.8 Transfection

Transfection of *P. falciparum* was carried out on synchronised cultures (section 2.9.5) with 4 – 7% ring stages (Crabb and Cowman, 1996; Crabb *et al.*, 2004). The medium was removed from the culture and the remaining erythrocytes were resuspended in cytomix (section 2.3.5) and centrifuged at 400 g for 5 min at 4ºC. The supernatant was discarded. A total of 30 µl of plasmid DNA (100 µg) in sterile TE buffer was carefully mixed with 370 µl of cytomix and 200 µl of the infected erythrocytes. Electroporation was carried out with a Gene Pulser Xcell electroporator (Bio-Rad) in 0.2 cm electroporation cuvettes (Bio-Rad) at 310 V and 950 µF and measuring time constant values of 10 – 15 ms. The cells were then immediately transferred into 9 ml of pre-warmed (to 37ºC) complete RPMI 1640 medium with 300 µl uninfected erythrocytes. The culture medium was changed after 6 h and the drug WR99210 was added to the culture. All transfections were carried out in duplicate on two separate days to increase chances of a successful transfection. The medium in the cultures was changed every day for the first 8 days and then every second day until parasites appeared, which took 3 – 6 weeks. After this, the cultures were maintained daily and parasites were harvested, DNA extracted (section 2.9.7) and a Southern blot was carried out to confirm genotype of the transfectant parasites before starting drug selection for integration of plasmid (section 2.9.9).

2.9.9 Gene knockout in *P. falciparum*

Knockout of *PfE2* and *PfaE3* was attempted using the pCC-1 plasmid (section 2.5.2.4), which integrates into the genome by double cross-over recombination via the 5’ and 3’ end gene fragments cloned into the vector (Maier *et al.*, 2006, Table 2-3). pCC-1 contains the *human dehydrofolate reductase* (*hdhfr*) gene for positive selection of trasfected parasites with the drug WR99210. Integration of the plasmid into the gene is forced by negative selection via the *S. cerevisiae cytosine deaminase* bifunctional chimera with *S. cerevisiae* uracil phosphoribosyl transferase gene (*cyd*) found on the pCC-1 plasmid. This gene converts 5-fluorocytosine (5-FC) to 5-fluorouracil, which inhibits RNA synthesis and thymidylate synthase. Thus parasites with episomal pCC-1 plasmid will be killed. If double cross-over recombination of the plasmid occurs, the *cyd* gene is lost and only parasites with integration of the positive selectable marker should survive in the culture.

Following confirmation of successful transfection of parasites by Southern blotting (section 2.5.11), Ancotil® (commercial name for 5-FC) was added to a final concentration of 1 μM in addition to WR99210 (at 5 nM final concentration). 5-FC killed most parasites in the first few days, however, after 1-2 weeks the parasites re-emerged. Parasites were harvested and DNA was extracted (section 2.9.7) and analysed for integration of plasmid by Southern blotting.

2.9.10 Cloning by limiting dilution

To obtain clonal populations of parasites that had correct integration of the pCC-1 plasmid, cloning by limiting dilution was carried out. This involved diluting the culture in a 96-well microtitre plate to 0.25 parasites/well with medium containing appropriate drugs. The medium was changed every 5 days for 15 days after which the wells were checked for growth of parasites every 2-3 days by placing a drop of resuspended erythocytes onto a slide, staining it with Giemsa and examining it with a microscope (section 2.9.2). Once parasite growth was confirmed the culture volumes were increased and parasites were harvested and DNA was extracted (section 2.9.7) and analysed to determine the genotypes of the clones by Southern blotting (section 2.5.11).

2.9.11 Growth assay

To investigate any effect gene knockout had on parasite growth compared with the wild type, a growth assay was carried out as described by Sanders *et al*. (2006) with modifications. Synchronised cultures of rings (section 2.9.5) were diluted to 0.5% parasitemia and 5% haematocrit in 2 ml of medium. Each parasite line was assessed in triplicate. Every second day the cultures were diluted 1:5 with fresh erythrocytes. Blood smears of the cultures were made every day at the same time for 14 days. The parasitemia for each day was determined by counting the infected erythrocytes in 1000 cells (section 2.9.2). Also, the cycle stage of the parasite lines was determined by counting 100 parasites and noting the presence of rings, trophozoites and schizonts.

2.9.12 Determination of IC_{50} values

The incorporation of [^{3}H]-hypoxanthine was used to determine IC_{50} values (Desjardins *et al.*, 1979) for L-buthionine sulfoximine (BSO), paraquat and triclosan. The starting concentrations for the agents were 1.25 mM for BSO, 0.5 mM for paraquat and 0.4 mM for tricolsan. Synchronised ring stage cultures (section 2.9.5) were diluted to 0.5% parasitemia, 2% haematocrit, in IC_{50} medium (section 2.3.5). A serial dilution of the drug at 2 times the concentration to be tested was carried out in a 96-well microtitre plate leaving 100 µl in each well. On addition of 100 µl of the parasites, the final drug concentration was obtained. Uninfected erythrocytes and parasites incubated without the drug were included as controls. The microtitre plates were placed into a 5% CO_2 incubator at 37ºC for 48 h. After 48 h, 100 µl of medium was removed and 100 µl of fresh medium with 5 µCi [^{3}H]-hypoxanthine/ml was added. The plates were incubated for a further 24 h after which they were frozen at -20ºC. The wells were defrosted at room temperature for 2-3 h before harvesting with a Harvester 96™ Mach III (TomTec) onto Printed Filter Mat A filter mats (Perkin Elmer). The filter mats were dried at 55ºC for 90 min and sealed into plastic sample bags. Following the addition of 4 ml of scintillation fluid, the decay was measured with a Wallac 1450 MicroBeta Trilux liquid scintillation counter (Perkin Elmer) for 1 min per well. IC_{50} values were calculated using the 'log (inhibitor) versus response' model in GraphPad Prism 5.0.

Chapter 3

Expression and purification of recombinant *P. falciparum* dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase

3.1 Dihydrolipoamide acetyltransferase

3.1.1 Introduction

Dihydrolipoamide acetyltransferase (E2) (EC 2.3.1.12) is an enzyme in the family of acyltransferases, which transfer non-amino-acyl groups (EC 2.3.1.-) using coenzyme A (CoA) as an acceptor. E2 catalyses the transfer of an acetyl group bound to the E1 cofactor thiamine pyrophosphate (TPP) to CoA, generating acetyl-CoA (Fig. 3-1).

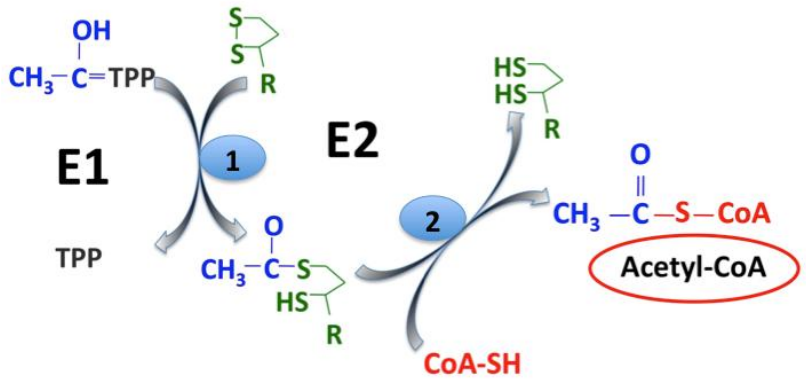


Figure 3-1. Reaction catalysed by E2
(1) The acetyl group (blue) is transferred from E1 via the thiamine pyrophosphate (TPP) cofactor to the lipoamide moiety of E2 (green), forming 8-acetyl-dihydrolipoamide. (2) From here the acetyl group is further transferred onto CoA (red) to form acetyl-CoA (circled), in the process reducing the lipoamide moiety.

P. falciparum E2 (*Pf*E2) (PlasmoDB annotation PF3D7_1020800, www.plasmodb.org) is located on chromosome 10 of the *P. falciparum* genome. The gene is 3192 bp in length, contains 10 introns and 11 exons. Complementary DNA analysis shows the full-length protein is 1923 bp and consists of 640 amino acids, with an estimated molecular mass of 73.9 kDa (McMillan, 2006). As the protein is located in the apicoplast (McMillan *et al.*, 2005), a signalling and transit peptide are present at the N-terminus of the protein. Using

subcellular localisation prediction programmes and alignments with E2 amino acid sequences from other organisms, the predicted start site of the mature *Pf*E2 was estimated at Ser51 resulting in a 67.2 kDa protein (McMillan, 2006).

The mature *Pf*E2 contains two lipoyl-domains (LD), an E1/E3 subunit binding domain (SBD), and a catalytic domain (CD) separated by long flexible linkers (Fig. 1-9). The structure as well as the stochiometry of the *Pf*E2 core are unknown. In humans, Gram-positive bacteria and plant mitochondria E2 forms an icosahedral 60-mer, while in Gram-negative bacteria the core is an octahedral 24-mer (Mattevi *et al.*, 1992a; Izard *et al.*, 1999; Mooney *et al.*, 1999; Yu *et al.*, 2008b). As the apicoplast is of red algal origin (Botte *et al.*, 2011), the hypothesis tested here was whether the *Pf*E2 forms a 24-mer, which is significantly different to the human E2 oligomeric structure and may have implications in the drug discovery process.

3.1.2 Aims

The aims for this part of the project were to:

- express recombinant *P. falciparum* E2 (r*Pf*E2) in soluble form
- purify r*Pf*E2
- determine the kinetic parameters of r*Pf*E2

3.1.3 Results

3.1.3.1 E2 amino acid alignments

Amino acid alignments were performed using ClustalW2 (Larkin *et al.*, 2007; Goujon *et al.* 2010). Jalview Version 2.8 (Waterhouse *et al*., 2009) was used to colour different groups of residues in the alignments. The N-terminal section of the proteins aligned poorly owing to the variation in number of lipoyl domains (LD) present in an organism. *Homo sapiens* and *P. falciparum* E2 and *Arabidopsis thaliana* mitochondrial E2 have two LDs, *A. thaliana* chloroplast and the cyanobacterium *Synechocystis* sp. PCC 6803 have one and *Azotobacter vinelandii* and *E. coli* E2 have three. Also the amino acids in the linkers vary greatly. Thus, to identify the conserved residues in the LDs, the sequences of those domains were aligned separately (Fig. 3-2).

P. falciparum LD1 shared greatest sequence identity with the single *Synechocystis* sp. LD (51%) followed by the single *A. thaliana* chloroplast LD (48%) and shared higher identity with the eukaryotic domains (28-34%) than the bacterial domains (19-29%). *P. falciparum* LD2, however, shared little sequence identity with other LDs (21-28%). It had greatest sequence identity with the single *Synechocystis* sp. LD (28%) and the single *A. thaliana* chloroplast LD and mitochondrial LD1 (both 26%). All the domains contain the lysine residue required for lipoylation (in *P. falciparum* these are Lys93 and Lys223, respectively, all numberings based on the full-length *Pf*E2 amino acid sequence, Fig. 3-2). The lysines are in the middle of a conserved DKA motif forming a sharp β-turn in the domain structure essential for recognition of the lysine for lipoylation (Russell and Guest, 1991; Dardel *et al.*, 1993; Wallis and Perham, 1994). This motif is present in *P.falciparum* LD1, but in LD2 the sequence is DKS. The identity of the aspartic acid or alanine adjacent to the lysine is not crucial for lipoylation of the lysine as long as the β-turn is preserved, however, recognition of the domain by E1 could be affected (Wallis and Perham, 1994).

Figure 3-2. Amino acid sequence alignment of E2 lipoyl domains (LD) from various organisms

ClustalW2 alignments were coloured based on amino acid type using Jalview Version 2.8. Peach, aliphatic/hydrophobic residues (ILVAM); orange, aromatic residues (FWY); blue, positive residues (KRH); red, negative residues (DE); green, hydrophilic residues (STNQ); magenta, conformationally special residues (PG); bright yellow, cysteine (C). Highly conserved residues denoted with a star. Conserved motifs are boxed. At_mE2, *A. thaliana* mitochondrial E2; At_pE2, *A. thaliana* chloroplast E2; Av_E2, *A. vinelandii* E2; Ec_E2, *E. coli* E2; Gs_E2, *G. stearothermophilus* E2; Hu_E2, human E2; Pf_E2, *Pf*E2, Sy_E2, *Synechocystis* sp E2. Number following LD denotes the domain number starting from the N-terminus.

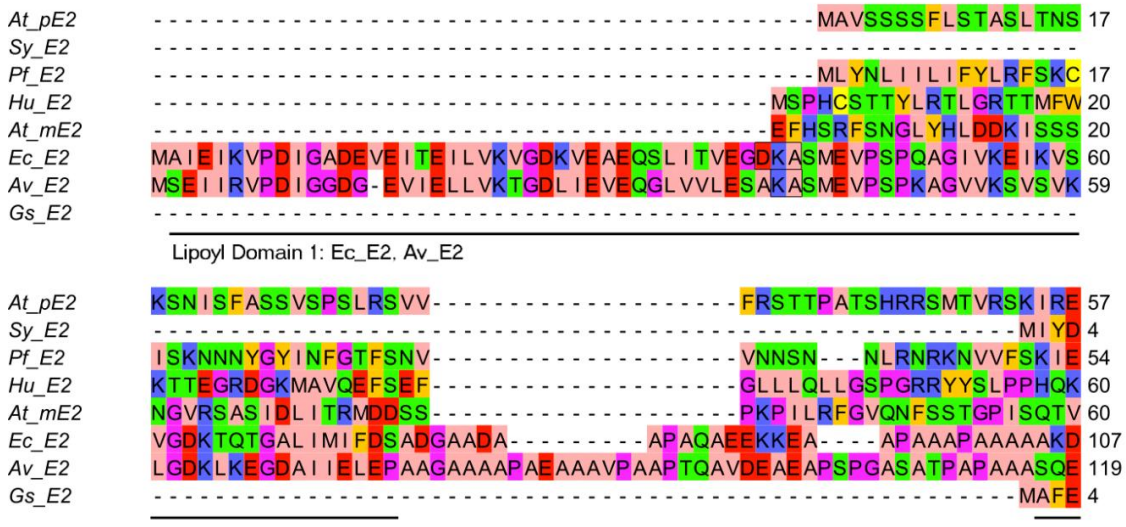
The eukaryotic E2 LDs, *Synechocystis* sp. and *P. falciparum* LD1 also contain the highly conserved P_{58}S/ALSP/STM_{64} sequence (the subscript number denotes the amino acid number based on the full-length *Pf*E2 amino acid sequence) (Neagle *et al.*, 1989). This motif is not present in the bacterial sequence and interestingly is not found in *P. falciparum* LD2. In *Pf*LD1, 15% of the residues were of a different amino acid type compared with the LDs from other organisms. The corresponding value for *Pf*LD2 was 30%, thus showing poorer conservation of this domain.

The amino acid sequence identity of mature length *Pf*E2 to other species was low, the highest being 34% with *Synechocystis* sp. followed by 29% identity with *A. thaliana* chloroplast E2 (Table 3-1).

Organism	Identity (%)
Synechocystis sp.	34
A. thaliana - chloroplast	29
G. stearothermophilus	22
A. thaliana - mitochondrion	22
H. sapiens	21
E. coli	19
A. vinelandii	18

Table 3-1. Amino acid identities of full-length *Pf*E2 compared with E2 from other organisms

The full-length E2 sequence alignments from various organisms are shown in two parts as the variation in linker lengths distorts alignment of the whole sequence. The LDs and linkers are aligned in Fig. 3-3 and the binding domain and catalytic domain in Fig. 3-4.



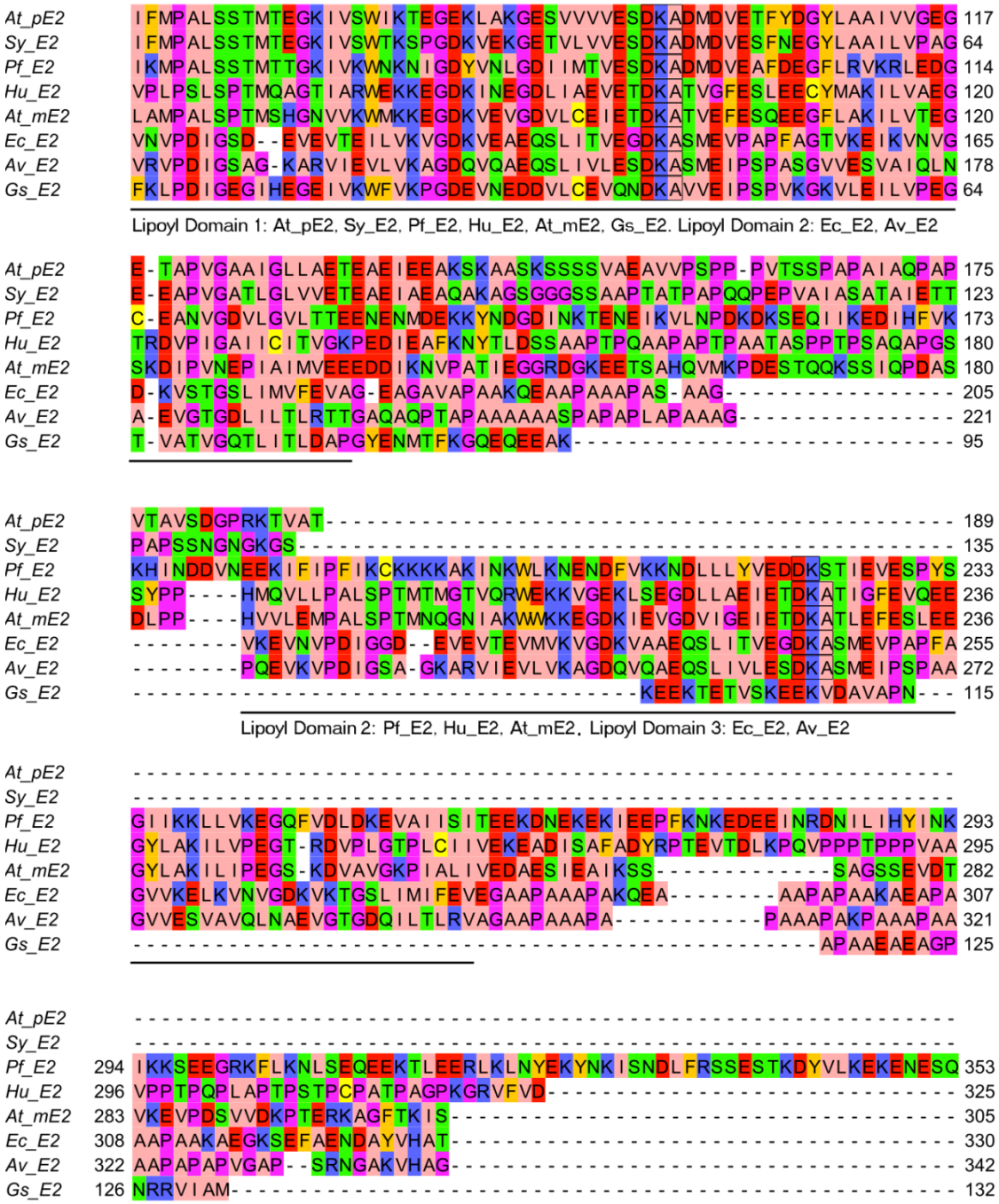


Figure 3-3. Amino acid sequence alignment of lipoyl domains and linkers from various organisms

ClustalW2 alignments were coloured based on amino acid type using Jalview Version 2.8. Peach, aliphatic/hydrophobic residues (ILVAM); orange, aromatic residues (FWY); blue, positive residues (KRH); red, negative residues (DE); green, hydrophilic residues (STNQ); magenta, conformationally special residues (PG); bright yellow, cysteine (C). Conserved motifs are boxed. At_mE2, *A. thaliana* mitochondrial E2; At_pE2, *A. thaliana* chloroplast E2; Av_E2, *A. vinelandii* E2; Ec_E2, *E. coli* E2; Gs_E2, *G. stearothermophilus* E2; Hu_E2, human E2; Pf_E2, *Pf*E2, Sy_E2, *Synechocystis* sp E2.

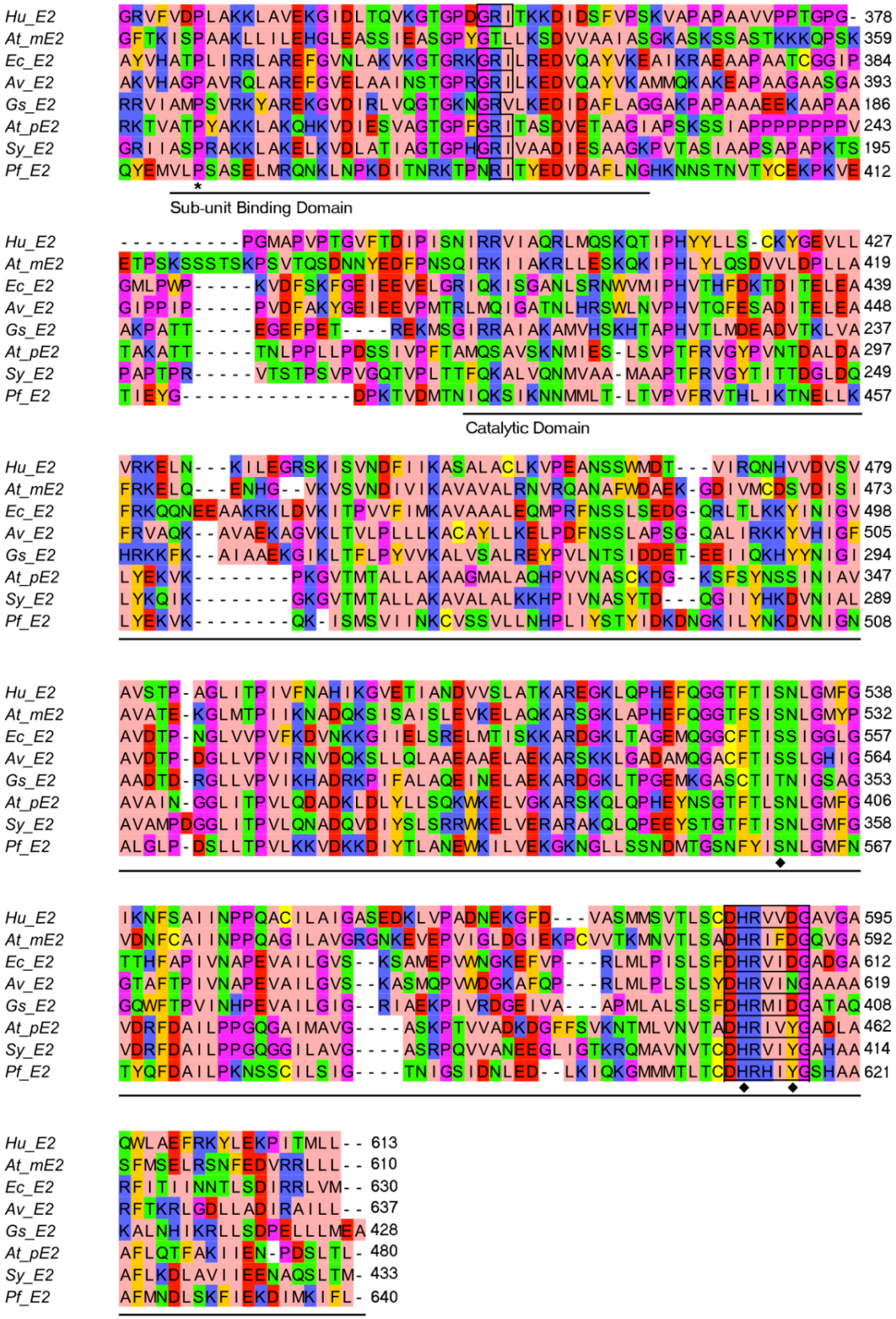


Figure 3-4. Amino acid sequence alignment of E2 binding domain and catalytic domain from various organisms

ClustalW2 alignments were coloured based on amino acid type using Jalview Version 2.8. Peach, aliphatic/hydrophobic residues (ILVAM); orange, aromatic residues (FWY); blue, positive residues (KRH); red, negative residues (DE); green, hydrophilic residues (STNQ); magenta,

conformationally special residues (PG); bright yellow, cysteine (C). Conserved proline in SBD denoted with a star. Conserved motifs are boxed. Active site residues denoted with a diamond shape. At_mE2, *A. thaliana* mitochondrial E2; At_pE2, *A. thaliana* chloroplast E2; Av_E2, *A. vinelandii* E2; Ec_E2, *E. coli* E2; Gs_E2, *G. stearothermophilus* E2; Hu_E2, human E2; Pf_E2, *Pf*E2, Sy_E2, *Synechocystis* sp E2.

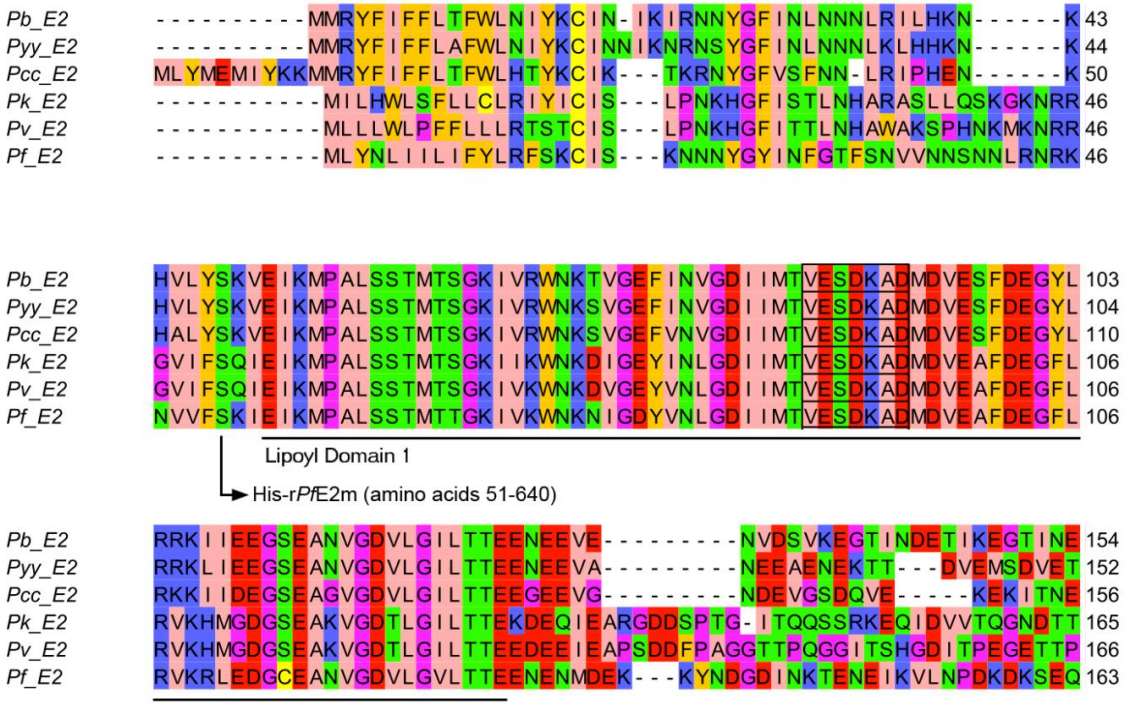
While human E2 is rich in small and hydrophobic amino acids (proline 10.3%, alanine 9.5%, leucine 8.0% and valine 8.0%), in *Pf*E2 the basic amino acid lysine is the most abundant (13.2%) followed by hydrophobic isoleucine (9.1%) and leucine (8.9%) and the amides glutamine (9.1%) and asparagine (8.9%). The amino acid profiles in the domains of *P. falciparum* and human E2 are similar with leucine (10.6% and 9.1% in *P. falciparum* and human domains, respectively), isoleucine (9.8% and 8.6%) and lysine (7.7% and 12.3%) being the most abundant residues. *Pf*E2 domains do contain, however, more asparagine (8.0% in *P. falciparum*, 2.9% in human domains). The starkest difference between the two proteins is found in the linkers, which in human E2 are high in proline (27.1%) and alanine (17.1%), whereas in the *Pf*E2 linkers, glutamine (18.3%), lysine (15.6%) and asparagine (11.8%) are the most abundant. The sizes of the domains are all similar, however, there are differences in linker lengths especially in the one between the second lipoyl domain and the binding domain. This linker is very long (96 amino acids) in *Pf*E2 compared with 40-59 amino acids in the other organisms.

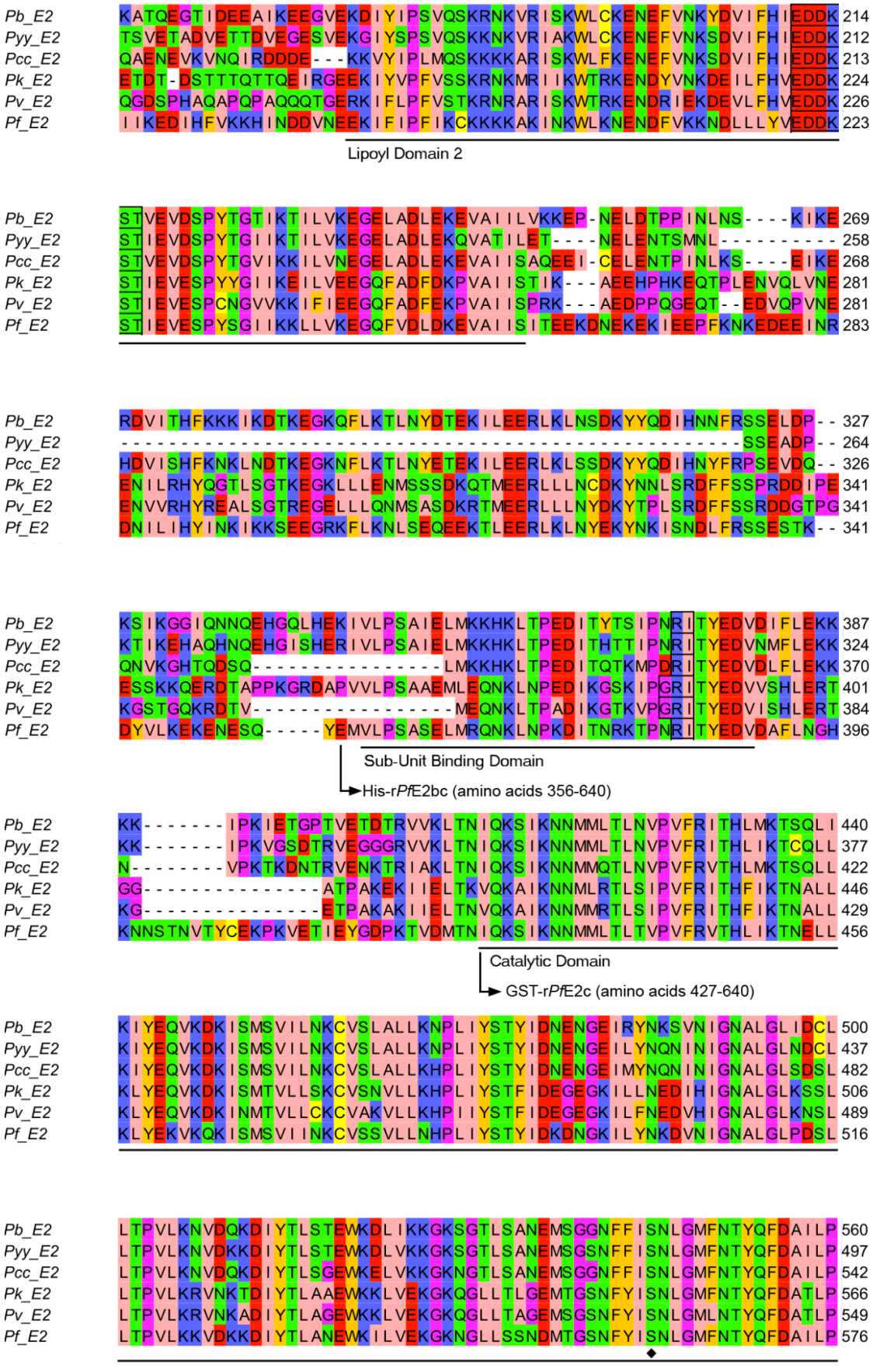
In the binding domain, important conserved residues include Pro359 (the numbering is based on the *P. falciparum* full-length sequence, Fig. 3-4), which promotes the formation of the first helix that interacts with E3 (Mande *et al.*, 1996). There is also a GRV/L/I domain thought to be important for the stability and function of the binding domain (Mande *et al.*, 1996), however, *P. falciparum* lacks the glycine and has an asparagine at this place instead (Asn382). This is unfavourable for the conformation of the domain, as a side chain in this position would protrude and could interfere with the recognition of the E3 binding residue (Arg383) by E3 (Kalia *et al.*, 1993). Also, the other residue involved in E3 binding is normally a positively charged arginine or lysine, but in *P. falciparum* this is an uncharged serine (Ser362).

In the catalytic domain there is a D_{611}HRXXD/N/YG_{617} motif (the subscript number denotes the amino acid number based on the full-length *Pf*E2 amino acid sequence, Fig. 3-4), which forms part of the active site and is conserved throughout species (Mooney *et al.*,

1999). The *P. falciparum* and *A. thaliana* chloroplast E2 have the same motif as *Synechocystis sp*. E2 (DHRXXYG), which is interesting as the apicoplast and chloroplast are of cyanobacterial ancestry. In *A. vinelandii* the motif is DHRXXNG and in the other organisms DHRXXDG. The active site of the catalytic domain in *P. falciparum* is formed by the residues Ser561, His612, Tyr616. The histidine is highly conserved and is present in all organisms and the Tyr616 varies as mentioned in the above motif. Apart from *G. stearothermophilus*, Ser561 is also conserved.

E2 sequences from other *Plasmodium* species were also aligned with ClustalW2 (Fig. 3-5). The non-human parasites, *P. berghei*, *P. yoelii* and *P. chabaudi*, aligned well as a group with identities ranging from 73-83%. The other two human infective *Plasmodium* species, *P. vivax* and *P. knowlesi*, also aligned well with each other, with 77% sequence identity. The two groups did not align well with each other with identities ranging between 49% to 51%. *Pf*E2 did not align as well with either group. The protein sequence had highest identity with E2 from *P. yoelii* (60%) followed by *P. berghei* and *P. chabaudi* (both 57%) (Table 3-2). However, this was only slightly higher than *Pf*E2 identity with *P. knowlesi* and *P. vivax* (both 54%).




Pb_E2 KATQEGTIDEEAIKEEGVEKDIYIPSVQSKRNKVRISKWLCKENEFVNKYDVIFHIEDDK 214
Pyy_E2 TSVETADVETTDVEGESVEKGIYSPSVQSKKNKVRIAKWLCKENEFVNKSDVIFHIEDDK 212
Pcc_E2 QAENEVKVNQIRDDDE---KKVYIPLMQSKKKKARISKWLFKENEFVNKSDVIFHIEDDK 213
Pk_E2 ETDT-DSTTTQTTQEIRGEEKIYVPFVSSKRNKMRIIKWTRKENDYVNKDEILFHVEDDK 224
Pv_E2 QGDSPHAQAPQPAQQQTGERKIFLPFVSTKRNRARISKWTRKENDRIEKDEVLFHVEDDK 226
Pf_E2 IIKEDIHFVKKHINDDVNEEKIFIPFIKCKKKKAKINKWLKNENDFVKKNDLLLYVEDDK 223
Lipoyl Domain 2
Pb_E2 STVEVDSPYTGTIKTILVKEGELADLEKEVAIILVKKEP-NELDTPPINLNS----KIKE 269
Pyy_E2 STIEVDSPYTGIIKTILVKEGELADLEKQVATILET----NELENTSMNL---------- 258
Pcc_E2 STVEVDSPYTGVIKKILVNEGELADLEKEVAIISAQEEI-CELENTPINLKS----EIKE 268
Pk_E2 STIEVESPYYGIIKEILVEEGQFADFDKPVAIISTIK---AEEHPHKEQTPLENVQLVNE 281
Pv_E2 STIEVESPCNGVVKKIFIEEGQFADFEKPVAIISPRK---AEDPPQGEQT--EDVQPVNE 281
Pf_E2 STIEVESPYSGIIKKLLVKEGQFVDLDKEVAIISITEEKDNEKEKIEEPFKNKEDEEINR 283
Pb_E2 RDVITHFKKKIKDTKEGKQFLKTLNYDTEKILEERLKLNSDKYYQDIHNNFRSSELDP--327
Pyy_E2 ----------SSEADP--264
Pcc_E2 HDVISHFKNKLNDTKEGKNFLKTLNYETEKILEERLKLSSDKYYQDIHNYFRPSEVDQ--326
Pk_E2 ENILRHYQGTLSGTKEGKLLLENMSSSDKQTMEERLLLNCDKYNNLSRDFFSSPRDDIPE341
Pv_E2 ENVVRHYREALSGTREGELLLQNMSASDKRTMEERLLLNYDKYTPLSRDFFSSRDDGTPG341
Pf_E2 DNILIHYINKIKKSEEGRKFLKNLSEQEEKTLEERLKLNYEKYNKISNDLFRSSESTK--341
Pb_E2 KSIKGGIQNNQEHGQLHEKIVLPSAIELMKKHKLTPEDITYTSIPNRITYEDVDIFLEKK 387
Pyy_E2 KTIKEHAQHNQEHGISHERIVLPSAIELMKKHKLTPEDITHTTIPNRITYEDVNMFLEKK 324
Pcc_E2 QNVKGHTQDSQ----------LMKKHKLTPEDITQTKMPDRITYEDVDLFLEKK 370
Pk_E2 ESSKKQERDTAPPKGRDAPVVLPSAAEMLEQNKLNPEDIKGSKIPGRITYEDVVSHLERT401
Pv_E2 KGSTGQKRDTV----------MEQNKLTPADIKGTKVPGRITYEDVISHLERT 384
Pf_E2 DYVLKEKENESQ-----YEMVLPSASELMRQNKLNPKDITNRKTPNRITYEDVDAFLNGH396
Sub-Unit Binding Domain
His-rPfE2bc (amino acids 356-640)
Pb_E2 KK-------IPKIETGPTVETDTRVVKLTNIQKSIKNNMMLTLNVPVFRITHLMKTSQLI 440
Pyy_E2 KK-------IPKVGSDTRVEGGGRVVKLTNIQKSIKNNMMLTLNVPVFRITHLIKTCQLL 377
Pcc_E2 N--------VPKTKDNTRVENKTRIAKLTNIQKSIKNNMMQTLNVPVFRVTHLMKTSQLL 422
Pk_E2 GG----------ATPAKEKIIELTKVQKAIKNNMLRTLSIPVFRITHFIKTNALL 446
Pv_E2 KG----------ETPAKAKIIELTNVQKAIKNNMMRTLSIPVFRITHFIKTNALL 429
Pf_E2 KNNSTNVTYCEKPKVETIEYGDPKTVDMTNIQKSIKNNMMLTLTVPVFRVTHLIKTNELL456
Catalytic Domain
GST-rPfE2c (amino acids 427-640)
Pb_E2 KIYEQVKDKISMSVILNKCVSLALLKNPLIYSTYIDNENGEIRYNKSVNIGNALGLIDCL 500
Pyy_E2 KIYEQVKDKISMSVILNKCVSLALLKNPLIYSTYIDNENGEILYNQNINIGNALGLNDCL 437
Pcc_E2 KIYEQVKDKISMSVILNKCVSLALLKHPLIYSTYIDNENGEIMYNQNINIGNALGLSDSL 482
Pk_E2 KLYEQVKDKISMTVLLSKCVSNVLLKHPLIYSTFIDEGEGKILLNEDIHIGNALGLKSSL 506
Pv_E2 KLYEQVKDKINMTVLLCKCVAKVLLKHPIIYSTFIDEGEGKILFNEDVHIGNALGLKNSL 489
Pf_E2 KLYEKVKQKISMSVIINKCVSSVLLNHPLIYSTYIDKDNGKILYNKDVNIGNALGLPDSL 516
Pb_E2 LTPVLKNVDQKDIYTLSTEWKDLIKKGKSGTLSANEMSGGNFFISNLGMFNTYQFDAILP 560
Pyy_E2 LTPVLKNVDKKDIYTLSTEWKDLVKKGKSGTLSANEMSGSNFFISNLGMFNTYQFDAILP 497
Pcc_E2 LTPVLKNVDQKDIYTLSGEWKELVKKGKNGTLSANEMSGGNFFISNLGMFNTYQFDAILP 542
Pk_E2 LTPVLKRVNKTDIYTLAAEWKKLVEKGKQGLLTLGEMTGSNFYISNLGMFNTYQFDATLP 566
Pv_E2 LTPVLKRVNKADIYTLAGEWKKLVEKGKQGLLTAGEMSGSNFYISNLGMLNTYQFDATLP 549
Pf_E2 LTPVLKKVDKKDIYTLANEWKILVEKGKNGLLSSNDMTGSNFYISNLGMFNTYQFDAILP 576

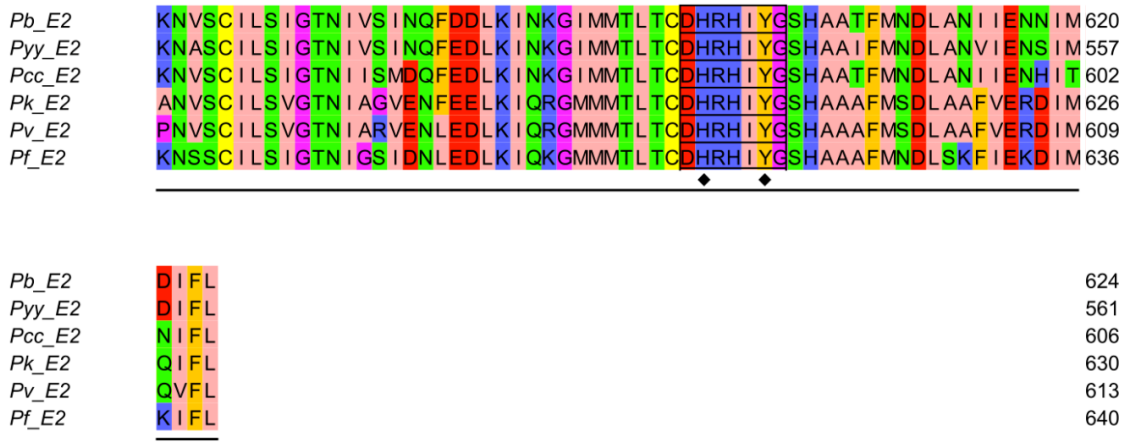


Figure 3-5. Amino acid alignment of E2 from various *Plasmodium* species
ClustalW2 alignments were coloured based on amino acid type using Jalview Version 2.8. Peach, aliphatic/hydrophobic residues (ILVAM); orange, aromatic residues (FWY); blue, positive residues (KRH); red, negative residues (DE); green, hydrophilic residues (STNQ); magenta, conformationally special residues (PG); bright yellow, cysteine (C). Conserved motifs are boxed. Active site residues denoted with a diamond shape. The start of each E2 construct used in this study is also shown. At_mE2, *A. thaliana* mitochondrial E2; At_pE2, *A. thaliana* chloroplast E2; Av_E2, *A. vinelandii* E2; Ec_E2, *E. coli* E2; Gs_E2, *G. stearothermophilus* E2; Hu_E2, human E2; Pf_E2, *Pf*E2, Sy_E2, *Synechocystis* sp E2.

Organism	Identity (%)
P. yoelii	60
P. berghei	57
P. chabaudi	57
P. knowlesi	54
P. vivax	54

Table 3-2. Amino acid identities of *Pf*E2 compared with PDC E2 from other *Plasmodium* species

When further examining the sequence alignments, when a residue was not conserved, the *Pf*E2 residue aligned 52% of the time with the non-human parasite sequence and 48% of the time with the human parasite sequence. Thus, *Pf*E2 seemed to contain a mix of residues from both groups. In addition, on 23 occasions when the residues of all the other species aligned, *Pf*E2 contained a different amino acid. This occurred less often in the other species (*P. yoelii n* = 11, *P. chabaudi n* = 6, *P. knowlesi n* = 6, *P. berghei n* = 1, *P. yoelii n* = 1).

While the domains are mostly highly preserved among all the *Plasmodium* species, differences can be seen in linker lengths. Most notably, *P. yoelii* has a very short linker between LD2 and the SBD (37 amino acids) compared with the other species (89-98 amino acids). Most of the conserved residues in each of the domains mentioned above are present in all the species. The asparagine instead of a glycine at Asn210 is conserved among the *Plasmodium* species. Interestingly, the binding domain is not as well conserved within the *Plasmodium* species as the domain is in other organisms.

3.1.3.2 Generation of expression constructs

On close inspection of the coding sequence for *Pf*E2, multiple rare codons were observed. These could hinder expression of the *Plasmodium* protein in *E. coli*, thus codon optimisation of the full-length *Pf*E2 with *E. coli* codon bias was carried out by GENEART AG (from 2010 onwards the company is known as GeneArt® Gene Synthesis, Life Technologies). Following gene synthesis, the company cloned both the full-length and the mature-length (starting at Ser51, Fig 3-5) *Pf*E2 sequences into the expression vector pET-15b, with a cleavable N-terminal 6 x His-tag, thus generating the constructs His-r*Pf*E2f and His-r*Pf*E2m, respectively.

The codon optimised *Pf*E2 sequence was also used as the template for generation of r*Pf*E2 constructs with only the binding domain and/or catalytic domain (see section 2.5.1.1, 2.5.2.1 and 2.5.2.2 for generation of constructs). Protein amino acid sequence alignments (Fig. 3-3 and 3-4) and assessment of homology models (section 4.1.3.1) were used to determine the start residue for each domain. One of these constructs, His-r*Pf*E2bc, consisted of amino acids 356-640 (Fig. 3-5), which contained the binding and catalytic domains and was cloned into pET-15b to include an N-terminal His-tag (this construct was generated by Jamie Whitelaw). The other, GST-r*Pf*E2c, included only the catalytic domain (amino acids 429-640, Fig. 3-5) and was cloned into pGEX-6P-1 with a GST-tag at the N-terminus.

3.1.3.3 Recombinant protein expression

To find the most favourable conditions for expression of the various recombinant proteins, small-scale expression trials were set up. By varying the expression vector, temperature, time and IPTG concentration the parameters required for optimum expression yield and purity of the recombinant protein were determined.

3.1.3.3.1 Expression of His-r*Pf*E2m

The initial trial to determine the optimal temperature for recombinant expression was carried out with His-r*Pf*E2f because this construct was obtained from GENEART AG prior to the His-r*Pf*E2m construct. The 76 kDa His-r*Pf*E2f in pET-15b was transformed into BLR(DE3) *E. coli*. This host strain is a BL21 derivative that can stabilise the expression of target proteins with repetitive sequences, which is often the case for *P. falciparum* genes. Expression was tested at various time points at 18°C, 25°C, 30°C and 37°C (Fig. 3-6).

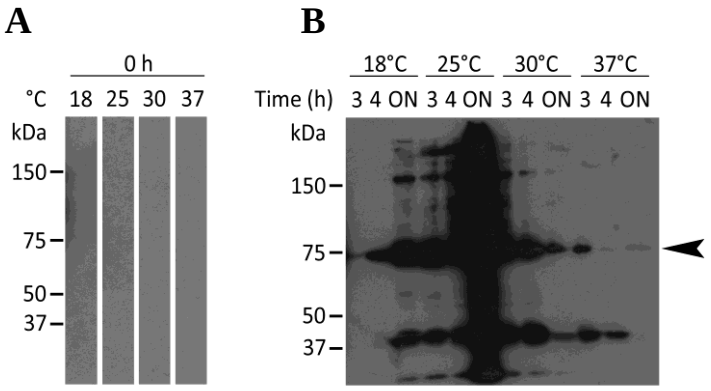


Figure 3-6. Western blot analysis of His-r*Pf*E2f expression in BLR(DE3) *E. coli*
Expression of r*Pf*E2f in pET-15b was tested at 18°C, 25°C, 30°C and 37°C following induction with 1 mM IPTG. Samples of expression cultures were taken (A) before induction and (B) 3 h, 4 h and overnight (ON) after induction. Expression of His-r*Pf*E2f was assessed by western blotting using an anti-His-tag antibody. Results for soluble fractions only are shown. Expression of His-r*Pf*E2f was tightly controlled as no recombinant protein was seen before induction. Recombinant, His-tagged protein of the expected size (76 kDa, arrow) was present on the blot with additional higher and lower molecular mass possibly corresponding to multimers of His-r*Pf*E2f or degradation products of the recombinant protein, respectively.

At 30°C and 37°C very little or no soluble His-r*Pf*E2f was detected, while at 25°C protein yield was high, however there was a lot of degradation and aggregation of the His-tagged protein. Expression at a lower temperature (18°C) for a shorter time (4 h) resulted in less degradation of the recombinant protein but a low yield. To avoid protein degradation, low temperatures were used in subsequent trials while attempting to improve yield.

His-r*Pf*E2m in pET-15b was transformed into BLR(DE3) *E. coli* and an expression trial with varying amounts of IPTG was carried out at 15°C and 18°C (Fig. 3-7). Slightly more protein was expressed at 18°C than at 15°C, probably due to differences in bacterial densities. Initially more degradation could be seen at 18°C, however, the difference was negligible after 2 h expression. The optimal IPTG concentration was deduced to be 0.5 mM. Increasing the concentration of IPTG did not result in higher levels of recombinant protein, but appeared to result in a greater level of degradation. The leaky expression seen in the non-induced controls after 3 h expression was unexpected as pET-15b contains the T7*lac* promoter, which should efficiently prevent transcription by basal levels of T7 RNA polymerase. However, as His-r*Pf*E2m was not toxic to the bacteria, the leaky expression was not such a concern. In all, expression of His-r*Pf*E2m in BLR(DE3) *E. coli* was poor, as degradation could already be seen 1 h after induction.

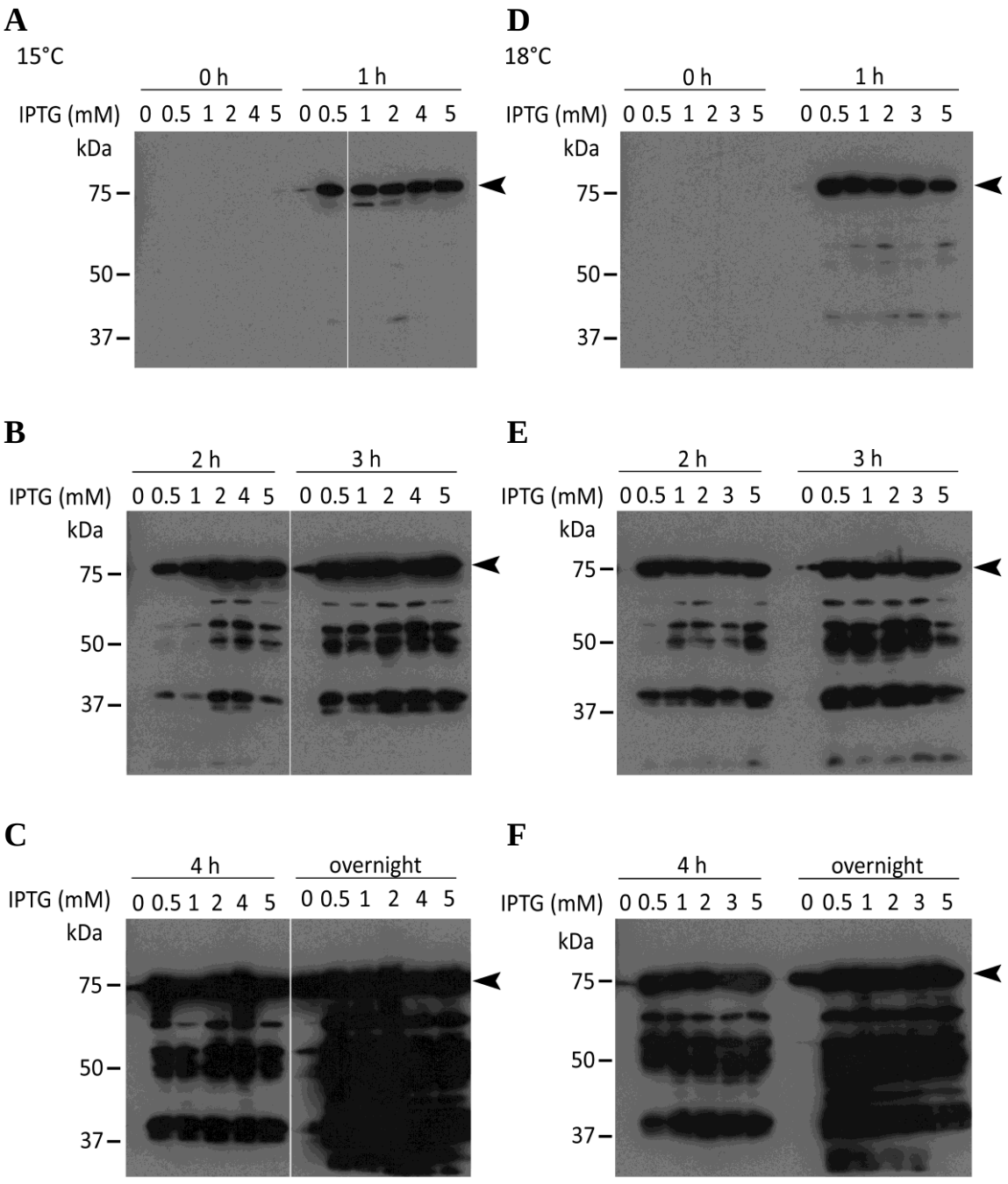


Figure 3-7. Western blot analysis of His-r*Pf*E2m expression in BLR(DE3) *E. coli*
The effect of various concentrations of IPTG (0 – 5 mM) on the expression of His-r*Pf*E2m in pET-15b was analysed at (A-C) 15°C and (D-F) 18°C. Samples of expression cultures were taken prior to induction (0 h) as well as after 1 – 4 h and overnight (ON) expression. Anti-His-tag antibody western blotting was used to assess the results; only soluble fractions are shown. His-r*Pf*E2m resolved on the gel as a ~75 kDa protein (arrow) although the estimated molecular mass from the amino acid sequence is 70 kDa. No expression of His-r*Pf*E2m was seen before induction (0 h). However, some leaky expression could be observed in the non-induced controls (0 mM IPTG) from 3 h onwards. Increasing the concentration of IPTG past 0.5 mM only resulted in increased amounts of degradation products.

Although His-r*Pf*E2m has an molecular mass of 70 kDa, the protein seemed to run at a larger 75 kDa size (Fig. 3-7), similar to r*Pf*E2f-His (76 kDa, Fig. 3-6). However, when the two proteins were electrophoresed in adjacent wells in the same gel the slight size difference between the two was observed (Fig. 3-8).

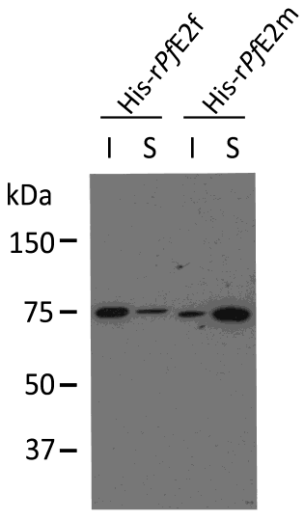


Figure 3-8. Size difference between His-r*Pf*E2f and His-r*Pf*E2m
Samples taken after 1 h expression showed a small size difference between both insoluble (I) and soluble (S) fractions of r*Pf*E2f-His (76 kDa) and His-r*Pf*E2m (70 kDa) when resolved on the same gel.

An expression trial of His-r*Pf*E2m in pET-15b was also carried out in Rosetta™ 2 (DE3)pLysS *E. coli* at 18°C (Fig. 3-9). This host *E. coli* contains additional plasmids encoding rare tRNAs. Expression of soluble His-r*Pf*E2m without degradation products was seen up to 2 h, with degradation and aggregation increasing from 3 h onwards. Nevertheless, compared with BLR(DE3) (Fig. 3-7), more recombinant protein and less degradation was observed.

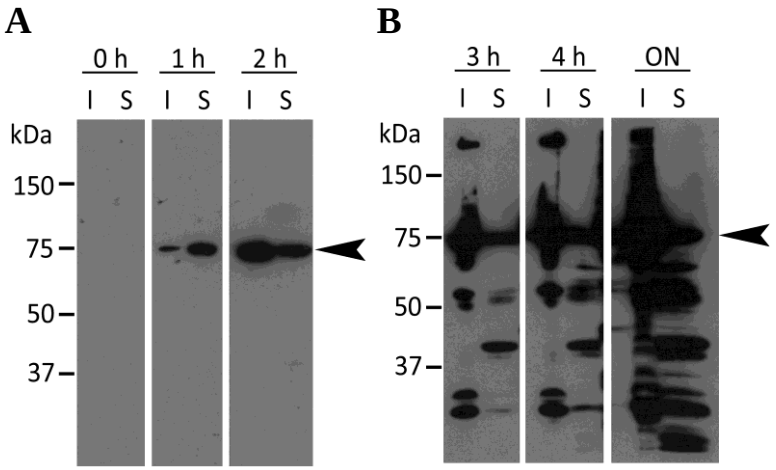


Figure 3-9. Western blot analysis of His-r*Pf*E2m expression in Rosetta™ 2(DE3)pLysS *E. coli*
Expression of His-r*Pf*E2m in pET-15b in Rosetta™ 2(DE3)pLysS *E. coli* was tested at 18°C following induction with 0.5 mM IPTG. Samples of expression cultures were taken before induction (0 h) and after 1 - 4 h and overnight (ON) expression (A & B). Anti-His-tag antibody western blotting was used to analyse results. His-r*Pf*E2m (arrow) was expressed in higher amounts in the insoluble fractions (I) compared with the soluble fractions (S). Degradation of His-r*Pf*E2m can again be seen as expression time increases.

In an attempt to decrease degradation of the recombinant protein, an expression trial was also carried out using BL21 Star™(DE3) *E. coli*. This line contains a mutated RNase E that cannot degrade mRNA, thus increasing its stability, hence the amount of full-length protein expressed should increase. The expression was, however, no better than with Rosetta™ 2(DE3)pLysS (Fig. 3-10).

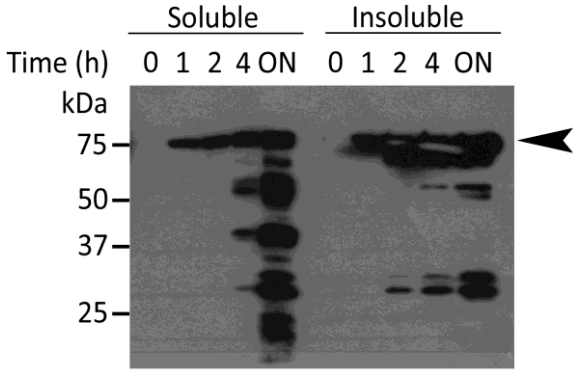


Figure 3-10. Western blot analysis of His-r*Pf*E2m expression BL21 Star™(DE3) *E. coli*
Expression was tested at 18°C following induction with 0.5 mM IPTG. Samples of expression cultures were taken before induction (0 h) and after 1 h – 4 h and overnight (ON) expression. His-r*Pf*E2m (arrow) was detected by an anti-His-tag antibody western blot. Results were comparable with expression in Rosetta™ 2(DE3)pLysS *E. coli*.

Based on all the trials Rosetta™ 2(DE3)pLysS was chosen as the bacterial line for large-scale expression of His-r*Pf*E2m. Cultures were grown in Terriffic broth (TB), rather than Luria Bertani (LB) broth, as it allowed growth of bacteria to higher densities ($OD_{600\ nm}$ = 0.6 - 0.8) during the log phase before induction of protein expression and subsequently an increased amount of recombinant protein was obtained. Initially expression for only 2.5 h was chosen to avoid degradation and aggregation of His-r*Pf*E2m. Following very low final yields of purified recombinant protein, higher yield of His-r*Pf*E2m was chosen over degradation, and overnight expression of His-r*Pf*E2m was carried out prior to purification of the protein used for all subsequent analyses.

3.1.3.3.2 Expression of His-r*Pf*E2bc

His-r*Pf*E2bc in pET-15b was transformed into BLR(DE3) and Rosetta™ 2(DE3)pLysS *E. coli*. No expression of soluble His-r*Pf*E2bc was seen in Rosetta™ 2(DE3)pLysS *E. coli*. Expression using BLR(DE3) resulted in low levels of soluble protein with no degradation at temperatures under 30°C (Fig. 3-11). For large-scale purification, His-r*Pf*E2bc was expressed for 18 h at 25°C in Terrific broth.

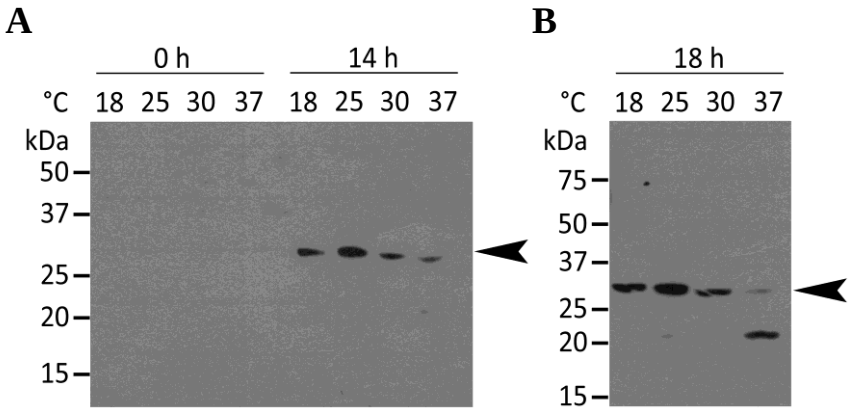


Figure 3-11. Western blot analysis of expression of His-r*Pf*E2bc in pET-15b
An expression trial of His-r*Pf*E2bc in BLR(DE3) was carried out at 18°C, 25°C, 30°C and 37°C. Samples of expression cultures were taken (A) before induction (0 h), 14 h and (B) 18 h after induction of expression with 1 mM IPTG. The soluble fractions were analysed by anti-His-tag antibody western blotting. Expression of the 35 kDa r*Pf*E2bc (arrow) was well controlled as no recombinant protein was seen at 0 h. Overall expression yield was low but largely free of degradation products. At 37°C r*Pf*E2bc degraded.

3.1.3.3.3 Expression of GST-r*Pf*E2c

GST-r*Pf*E2c was transformed into Rosetta™ 2(DE3)pLysS and BLR(DE3) *E. coli*. Expression trials at various temperatures (18°C, 25°C, 30°C and 37°C) resulted in very degraded expression in Rosetta™ 2(DE3)pLysS (Fig. 3-12A & B). Successful expression was achieved in BLR(DE3), in which overnight expression at 18°C resulted in the highest amounts of soluble GST-r*Pf*E2c with minimal degradation (Fig. 3-12C-D). These conditions were chosen for large-scale expression.

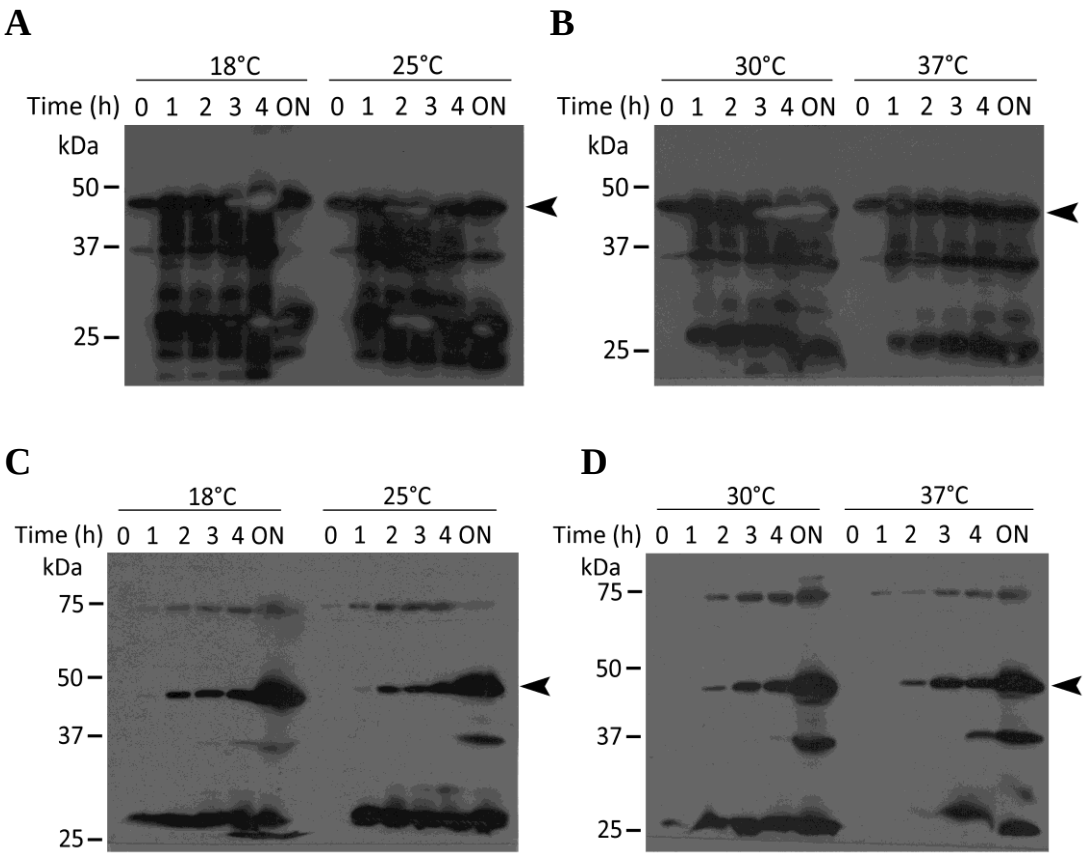


Figure 3-12. Western blot analysis of expression of GST-r*Pf*E2c in pET-15b
Expression of GST-r*Pf*E2c was trialled in two bacterial lines at 18°C, 25°C, 30°C and 37°C. Samples of expression cultures were taken before induction (0 h) and 1 – 4 h and overnight after induction with 1 mM IPTG. Soluble fractions were analysed by anti-His-tag antibody western blotting. (A & B) Expression in Rosetta™ 2(DE3)pLysS was leaky as some protein can be seen in the 0 h sample and large amounts of degradation can be seen. The GST-r*Pf*E2c was expected to be 52 kDa, however, it resolved as a lower molecular mass band at about 45 kDa (arrow). (C & D) Expression in BLR(DE3) was much cleaner and tightly controlled as no protein can be seen at 0 h. Again the 52 kDa GST-r*Pf*E2c (arrow) resolved at a lower molecular mass.

3.1.3.4 Recombinant protein purification

3.1.3.4.1 Purification of His-r*Pf*E2m

Initial purification of His-r*Pf*E2m was carried out with nickel affinity chromatography using a 1 ml HisTrap FF column. Protein lysate from 4 L of His-r*Pf*E2m culture expressed at 18°C for 2.5 h was loaded onto the column. A preliminary gradient elution (10 mM – 1 M imidazole) of His-r*Pf*E2m was used to determine the elution profile of the protein and any contaminants. Based on this, a step-wise increase in imidazole concentration was used to separate His-r*Pf*E2m from contaminants eluting at concentrations up to 400 mM. His-r*Pf*E2m eluted in peak 3 and the fractions containing the highest amounts of the protein were pooled for further purification by gel filtration (Fig. 3-13A-C). Approximately 3 mg of His-r*Pf*E2m was loaded onto a HiLoad™ 16/60 Superdex 200 gel filtration column, which allows separation of proteins up to 600 kDa. His-r*Pf*E2m eluted as three species; a large molecular mass species in the void volume (≥600 kDa), and two others corresponding to 450 kDa and 170 kDa, respectively (Fig. 3-13D-G). Peak 4 (Fig. 3-13D) contained only very small molecular mass species (43 kDa), thus it was concluded not to be His-r*Pf*E2m and was not analysed further. Lipoylation of His-r*Pf*E2m was confirmed by western blot using anti-lipoic acid (α-LA) antibody (Fig 3-13H). Interestingly, while the void volume peak contained least His-r*Pf*E2m, this fraction showed the strongest α-LA signal. This suggested that the protein in this peak was highly lipoylated and could thus be the correctly folded and assembled His-r*Pf*E2m required for subsequent characterisation. However, the protein concentration was too low (0.02 – 0.03 mg/ml) for any structural analyses to be carried out successfully.

A major cause of concern was an additional slightly lower molecular mass protein seen in gel filtration peak 2 (Fig. 3-13F, grey arrow). As this was not detected by α-LA antibody, it was suspected to be an *E.coli* contaminant. Both protein bands in peak 2 and the band in peak 3 were excised and sent for identification using mass spectrometry (Proteomics Unit, University of Glasgow). Peak 3 was confirmed to be His-r*Pf*E2m, however, no good match was obtained for the 75 kDa protein in peak 2 and the smaller protein was identified as an *E.coli* bifunctional UDP-glucuronic acid decarboxylase/UDP-4-amino-4-deoxy-L-arabinose formyltransferase. Due to this contamination, purification using this method was not continued.

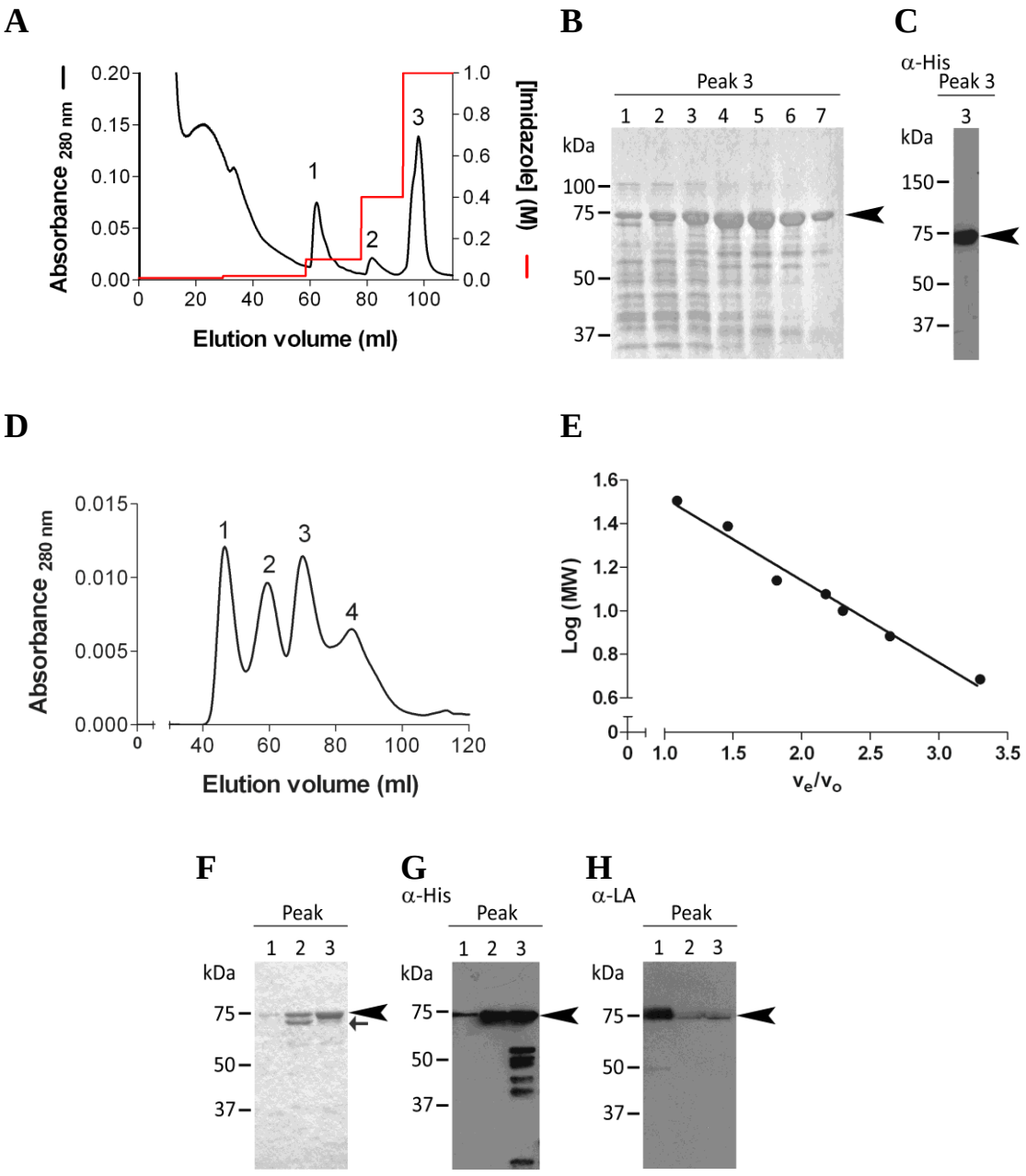


Figure 3-13. Purification of His-r*Pf*E2m with nickel affinity chromatography
(A) Protein lysate was loaded onto a 1 ml HisTrap FF column. A step-wise increase in imidazole concentration (100 mM, 400 mM and 1 M) was used to obtain a good separation of contaminating proteins (peak 1 and 2) from His-r*Pf*E2m (peak 3). (B) SDS-PAGE analysis shows a major protein of ~75 kDa (arrow) is present in peak 3. Numbers above lanes denote fractions. (C) Anti-His-tag antibody (α-His) western blot verified that the ~75 kDa protein (arrow) corresponds to His-r*Pf*E2m. (D) Fractions 3 – 6 of the HisTrap FF peak 3 were pooled and further purified by gel filtration using a HiLoad™ 16/60 Superdex 200 column. (E) A standard curve for the above column was obtained to determine the size of the protein present in each peak (see section 2.6.8.6 for standards). Elution peak 1 corresponded to the void volume, peaks 2, 3 and 4 eluted at volumes corresponding to 450 kDa, 170 kDa and 43 kDa respectively. (F) SDS-PAGE showed that the first three peaks contained the 75 kDa His-r*Pf*E2m (black arrow), but peak 2 also contained a slightly smaller size protein (grey arrow). (G) α-His-tag antibody western blot confirmed the 75 kDa protein as His-r*Pf*E2m (arrow). (H) Lipoylation of His-r*Pf*E2m was confirmed (arrow) when the western blot shown in figure G was stripped and probed with anti-LA antibody (α-LA).

Immobilised-metal affinity chromatography (IMAC) is based on the affinity of transition metal ions such as copper, nickel, zinc and cobalt to histidine and cysteine, and was first shown to be useful protein purification method in the 1970s (Block *et al.*, 2009). Following this, His-tags for recombinant proteins were developed to improve binding of target proteins to metal ions bound on agarose beads. However, repeats of histidine residues in native expression host proteins can also bind the metal ions leading to contaminating proteins in the purification fractions, as seen with the above nickel HisTrap FF purification. The metal used in the purification can be altered to increase selectivity for the recombinant protein. The commonly used metals, in order of binding strength, are copper, nickel, zinc and cobalt. Because the *E. coli* contaminant seen in peak 2 of gel filtration (Fig. 3-13F) was binding strongly to nickel, zinc, which does not bind His-tagged proteins as strongly, was chosen for further purification.

A POROS® PEEK™ metal affinity chromatography column loaded with $ZnSO_4$ was used to bind His-r*Pf*E2m from 6 L of culture expressed at 18°C for 2.5 h. His-r*Pf*E2m eluted in one major peak at ~800 mM imidazole when a linear gradient was applied (Fig 3-14A-C). The fractions with the highest protein concentration were pooled and ~1 mg was loaded onto a HiPrep 26/60 Sephacryl S-400 HR gel filtration column. This column has a fractionation range from 20 kDa to 8 MDa, thus it was chosen for better resolution of the larger molecular mass His-r*Pf*E2m species (≥600 kDa) that were present in the void volume of the HiLoad™ 16/60 Superdex 200 column (Fig. 3-13D). His-r*Pf*E2m eluted in two peaks, the first after 193 ml, the second at 222 ml and the protein in both peaks was lipoylated (Fig. 3-14D-G).

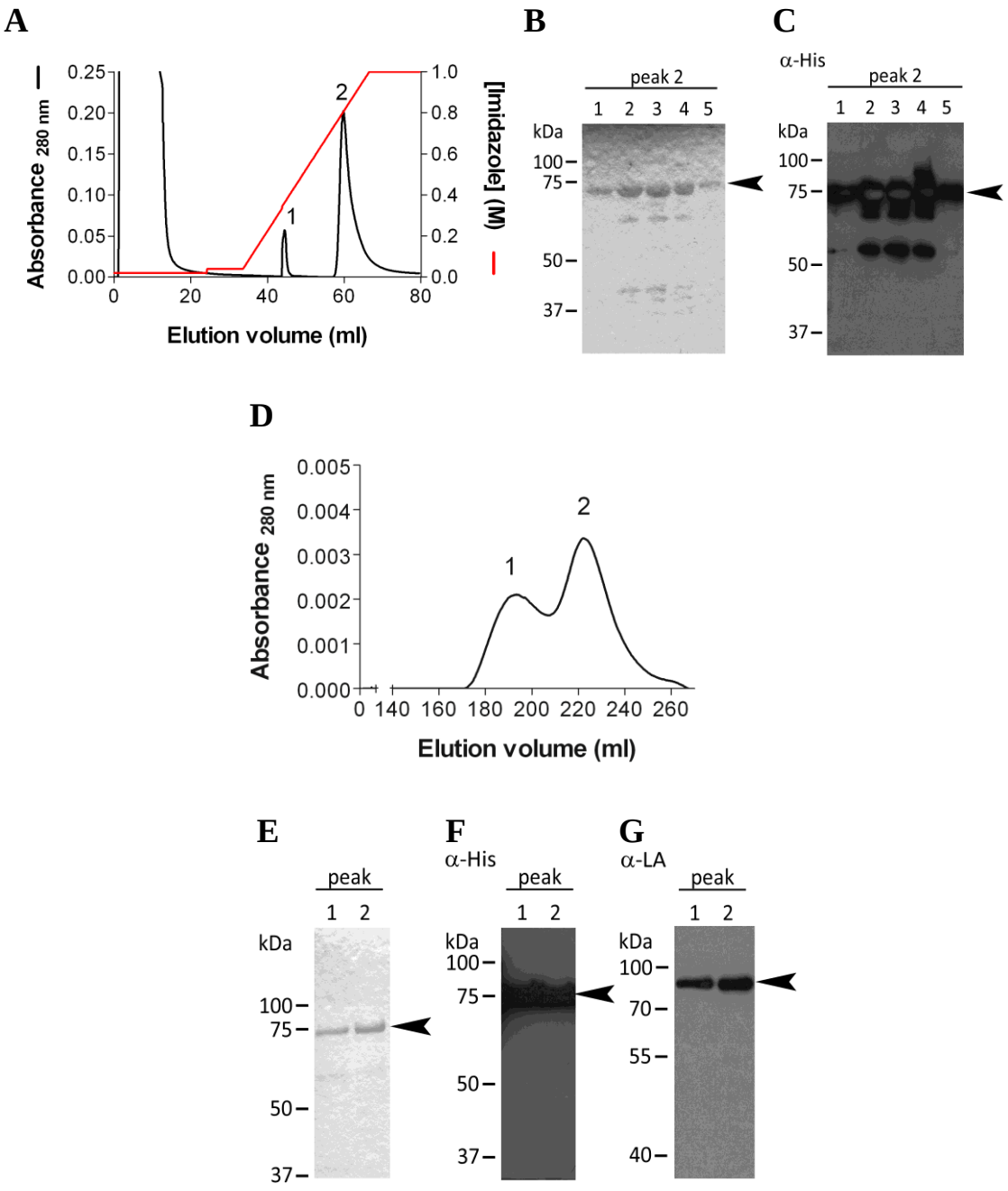


Figure 3-14. Zinc POROS® PEEK™ purification of His-r*Pf*E2m
A) A POROS® PEEK™ metal affinity chromatography column loaded with $ZnSO_4$ was used for further purification of His-r*Pf*E2m. A linear imidazole gradient (40 mM – 1 M) was used to elute His-r*Pf*E2m. (B) SDS-PAGE analysis shows the main protein present in peak 2 has a molecular mass of ~75 kDa (arrow). Numbers above lanes denote fractions. (C) Anti-His-tag antibody (α-His) western blot analysis confirmed this to be His-r*Pf*E2m (arrow) along with several degradation products. (D) Fractions 1 – 5 from the POROS® PEEK™ purification were pooled and further purified by gel filtration using a HiPrep 26/60 Sephacryl S-400 HR column. (E) A 75 kDa protein (arrow) was seen on SDS-PAGE of both gel filtration peaks. (F) The protein bands were verified to be His-r*Pf*E2m (arrow) with α-His-tag antibody western blot. (G) Both protein species were lipoylated as shown by an anti-lipoic acid antibody (α-LA) western blot.

Due to technical difficulties in determining the void volume for this particular gel filtration column, the molecular mass of the protein species cannot be determined from the chromatogram. Because the Sephacryl S-400 separates proteins up to 8 MDa, the typical high molecular mass standard, dextran blue (2 MDa), was not large enough to determine the exact void volume. The appropriate void marker for this particular column used by its manufacturer, GE Healthcare, was proprietary and not available for purchase. An attempt to determine the void volume was made using 0.5 μM latex beads, however, without a refractometer the elution of the beads could not be detected. Nevertheless, standards with known molecular masses were separated on the column. Dextran blue (2 MDa) eluted at 100 ml, thryoglobulin (669 kDa) eluted at 186 ml, apoferritin (440 kDa) at 196 ml, alcohol dehydrogenase (150 kDa) at 217 ml and bovine serum albumin (66 kDa) eluted at 240 ml. From this, the molecular masses for the protein species in gel filtration peaks 1 and 2 (Fig. 3-12D) could be estimated. Peak 1 eluted at 193 ml suggesting a molecular mass between 440 - 700 kDa. Peak 2 eluted at 222 ml, which could correspond to a molecular mass between 70 - 150 kDa.

Although His-r*Pf*E2m purified after 2.5 h expression was very pure, the protein concentration obtained was very low; only 10 ml of protein at ~20 μg/ml was obtained. Even after concentrating the protein extensively (from ~ 10 ml to 1ml), only up to 0.2 mg/ml of His-r*Pf*E2m could be obtained. Attempts to carry out analytical ultracentrifugation (AUC) to determine the oligomeric state of His-r*Pf*E2m at such a low protein concentration were unsuccessful. Thus, the decision was made to choose protein yield over degradation products and subsequent purifications were carried out on His-r*Pf*E2m expressed for a longer period to increase protein yield.

Protein lysate from 2 L of His-r*Pf*E2m culture expressed at 18°C overnight was used for purification. Due to the high viscosity of the lysate, this was divided into 5 aliquots, which were each purified using a $ZnSO_4$ POROS® PEEK™ column. Initially a 10 mM – 1 M imidazole gradient elution was carried out and His-r*Pf*E2m eluted as a doublet peak between 300 – 700 mM. To separate the two peaks, subsequent elutions were carried out using step-wise increments (Fig. 3-15A). Although on SDS-PAGE (Fig. 3-15B) both peaks seemed to contain the same proteins, only peak 1 was used in further purification as this contained most of the protein.

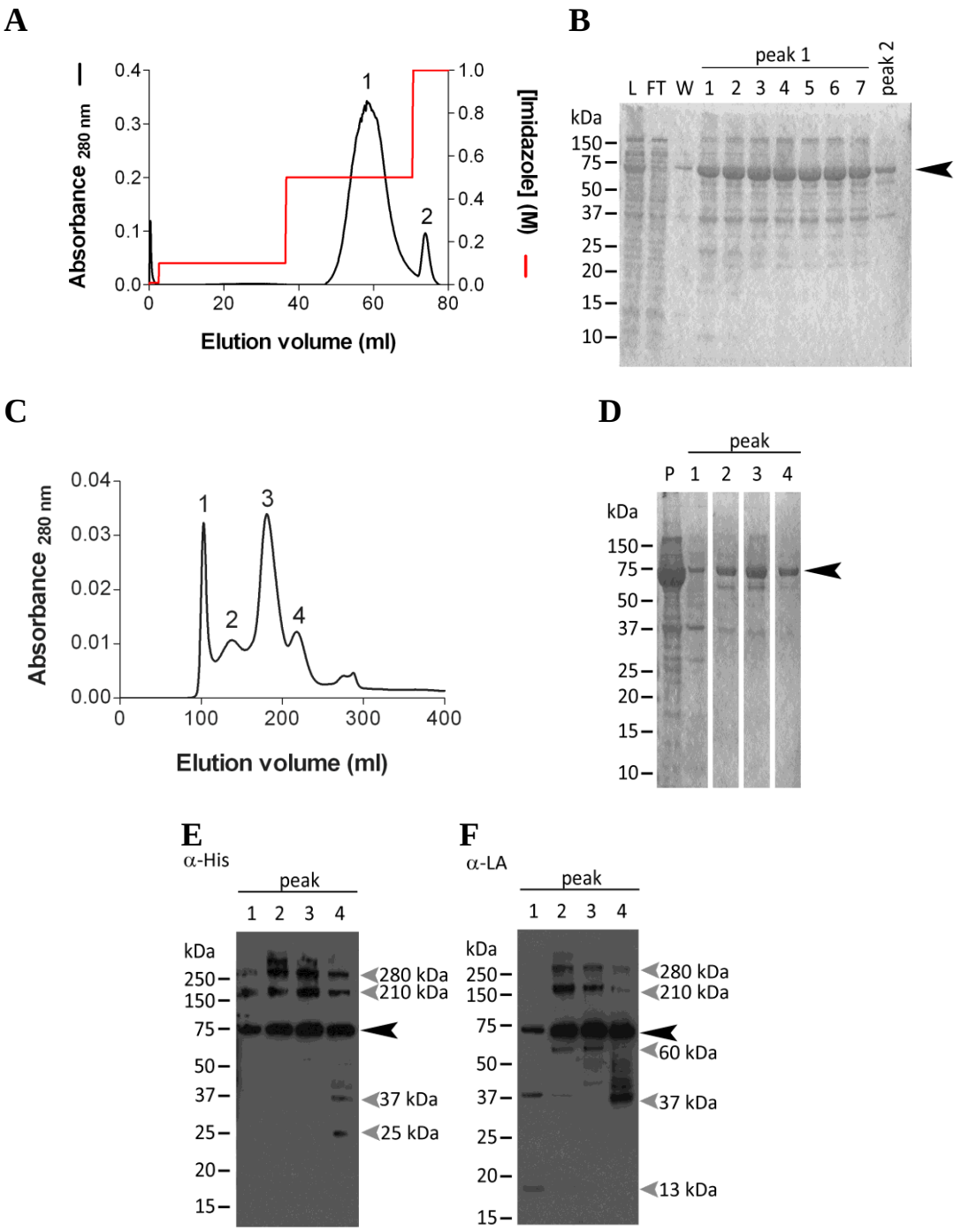


Figure 3-15. Purification of His-r*Pf*E2m following overnight expression
(A) Step-wise increase in imidazole concentration (100 mM, 500 mM and 1 M) was used to elute His-r*Pf*E2m from the $ZnSO_4$ POROS® PEEK™ column. (B) SDS-PAGE analysis showed the 75 kDa His-r*Pf*E2m (arrow) was mostly in peak 1. Numbers above lanes denote fractions. (C) Fractions 2 - 6 from the POROS® PEEK™ purification were pooled and purified by gel filtration using a HiPrep 26/60 Sephacryl S-400 HR column. (D) SDS-PAGE showed that His-r*Pf*E2m appears to be present in all four peaks. (E) Anti-His-tag antibody (α-His) western blot confirmed the presence of His-r*Pf*E2m (black arrow) but also showed degradation products (grey arrows, 25 and 37 kDa) and larger molecular mass bands (grey arrows, 210 and 280 kDa). (F) Anti-LA antibody (α-LA) western blot confirmed lipoylation of His-r*Pf*E2m (black arrow), the degradation products (grey arrows, 13, 37 and 60 kDa) and larger molecular mass bands (grey arrows, 210 and 280 kDa). L, protein lystate; FT, flow-through; W, wash; P, protein loaded onto gel filtration column.

At this point in the purification, the total protein obtained from 0.4 L of culture was 2.5 mg compared with 1 mg from 6 L His-r*Pf*E2m expressed for 2.5 h. The fractions with most His-r*Pf*E2m from five separate purifications were pooled, concentrated and loaded onto the HiPrep 26/60 Sephacryl S-400 HR gel filtration column. His-r*Pf*E2m eluted in four peaks, at 103 ml, 134 ml, 181 ml and 218 ml (Fig. 3-15C-D). As previously described, the molecular mass of each peak cannot be assigned accurately on the gel filtration column used. However, estimated molecular masses based on standards were ~ 2 MDa, >700 kDa, ~ 700 kDa and ~150 kDa, respectively for peaks 1-4. The protein concentrations were 0.08, 0.15, 0.64 and 0.25 mg/ml for each peak, respectively. For subsequent structural analyses His-r*Pf*E2m from each peak was concentrated using aquacide, as the protein bound onto the membrane on spin columns (Vivaspin™ concentrators).

Western blot analysis of His-r*Pf*E2m in each peak using α-His and α-LA antibodies (Fig 3-15E-F) showed protein bands larger than the monomer size, suggesting the presence of multimers, which were not disrupted in the process of denaturing the protein for the blots. The band sizes could correspond to a trimer (210 kDa) and possibly a tetramer (280 kDa) and even larger species (Fig. 3-15E-F). Several degradation products were also seen. A 60 kDa degradation product was detected by the anti-lipoic acid antibody but not the anti-His-tag antibody, suggesting the loss of the 2 kDa His-tag and part of the first lipoyl domain from this protein species. A second smaller product (~37 kDa) was detected by both the anti-lipoic and the anti-His-tag antibody, implying truncation of the protein nearer the catalytic domain. Finally a ~ 13 kDa band was detected only on the anti-lipoic acid western blot, which could correspond to a lipoyl domain in its own. Similar degradation products have been reported for full length human E2 (Yang *et al.*, 1997).

3.1.3.4.2 Purification of His-r*Pf*E2bc

Protein lysate from 2 L of His-r*Pf*E2bc culture expressed at 25°C for 18 h was loaded onto a 1 ml nickel HisTrap FF column. In the preliminary purifications, elution was carried out with a linear imidazole gradient first up to 1 M and then to 1.5 M, but His-r*Pf*E2bc did not elute within the gradient until the final concentration was increased to 2 M (Fig. 3-16A). His-r*Pf*E2bc was present mainly in the latter fractions of peak 3 (Fig. 3-16B&C), which eluted at ~ 1.2 M imidazole.

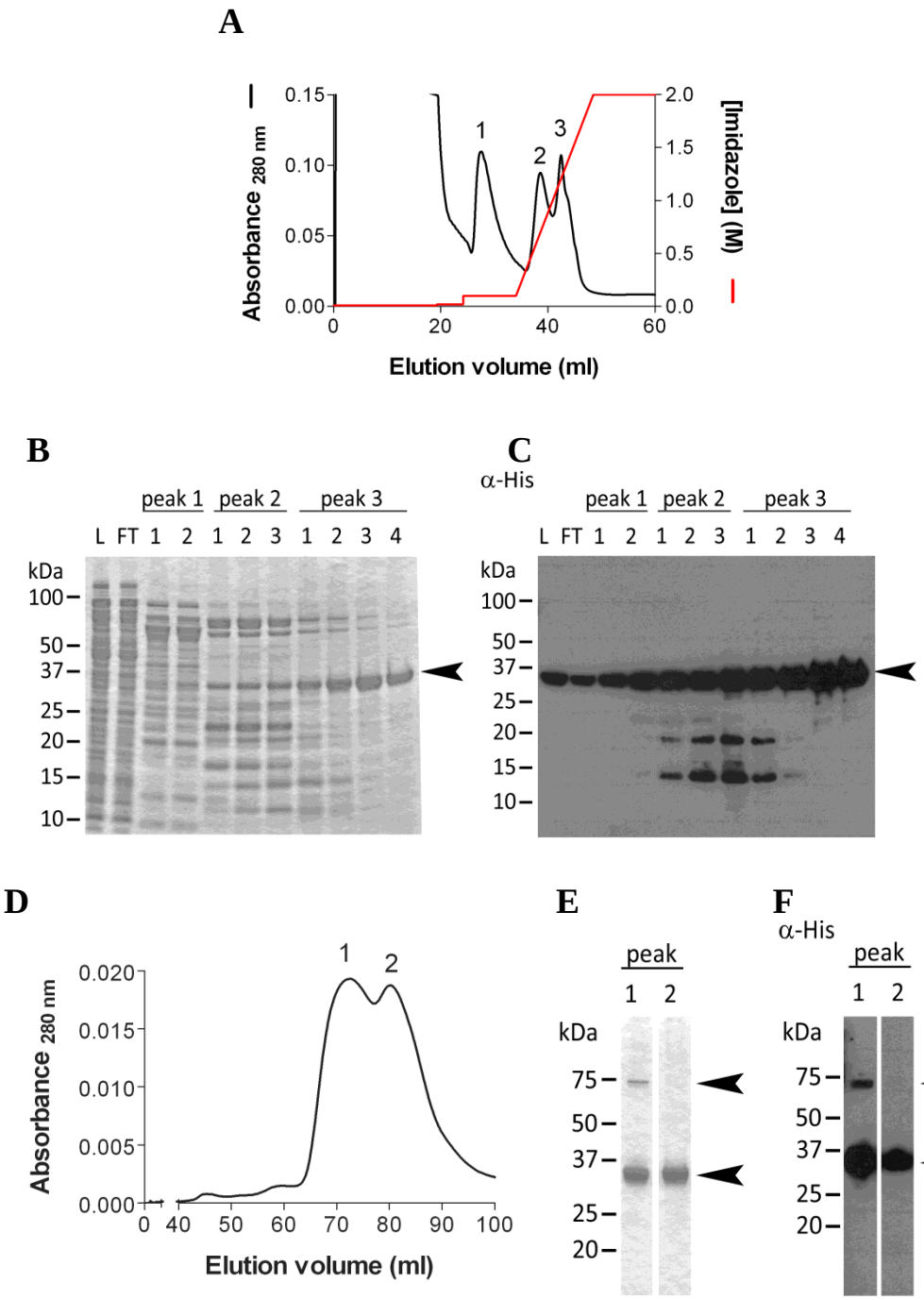


Figure 3-16. Purification of His-r*Pf*E2bc
(A) His-r*Pf*E2bc was eluted from the 1 ml HisTrap FF column using an imidazole gradient ranging from 100 mM to 2 M. (B) SDS-PAGE analysis showed that the largest amount of a 35 kDa protein (arrow) was purified in peak 3. (C) Anti-His-tag antibody western blot (α-His) analysis confirmed the 35 kDa band as His-r*Pf*E2bc (arrow) and showed fractions 2-4 of peak 3 to lack His-tagged degradation products. (D) Fractions 2 – 4 of peak 3 from seven separate purifications were pooled and further purified by gel filtration using a HiLoad™ 16/60 Superdex 200 column. His-r*Pf*E2bc eluted in two poorly resolved peaks corresponding to 140 kDa and 68 kDa, respectively. For standard curve see Fig. 3-13E. (E) SDS-PAGE shows peak 1 contained the 35 kDa protein (arrow) and an additional protein at around 75 kDa (arrow). (F) α-His western blot confirmed the presence of His-r*Pf*E2bc (arrow) and that the larger band (arrow) was also His-tagged. L, protein lystate; FT, flow-through

As these fractions were relatively free of contaminants, further optimisation of the elution was not carried out. Fractions from seven separate purifications were pooled, concentrated (~ 8 mg in 5 ml) and loaded onto the HiLoad™ 16/60 Superdex 200 gel filtration column. His-r*Pf*E2bc eluted in two peaks corresponding to molecular masses of 140 kDa and 68 kDa (Fig. 3-16D-F). While fractions in the second peak consisted of pure His-r*Pf*E2bc, SDS-PAGE showed the fractions in the first peak to contain a larger molecular mass protein. The protein was also detected by anti-His-tag antibody western blot, suggesting it was a multimer of His-r*Pf*E2bc that had not completely denatured during the SDS-PAGE protocol. The proteins in these two peaks were kept separate for subsequent analyses.

Peak 1 contained ~ 2 mg of His-r*Pf*E2bc in 11 fractions (11 ml). Peak 2 contained ~ 1 mg of His-r*Pf*E2bc in 10 fractions (10 ml). Concentration of the proteins was attempted using Vivaspin™ concentrators and aquacide. Unfortunately, His-r*Pf*E2bc was not stable at higher concentrations. This could already be seen during purification as ~ 8 mg of protein was loaded onto the gel filtration column but a total of only ~ 3 mg was collected in peak 1 and 2 elution fractions. The elution profiles showed broad peaks at 110 – 115 ml, corresponding to 3 – 4 kDa, which could be degraded His-r*Pf*E2bc. During concentration, precipitation of His-r*Pf*E2bc was seen once 0.5 – 0.7 mg/ml was reached. The addition of 2 mM EDTA and up to 5% glycerol or 5 mM DTT did not prevent aggregation or improve yield. Concentration of His-r*Pf*E2bc was also attempted by re-loading fractions from gel filtration onto a 1 ml nickel HisTrap FF column. The eluted His-r*Pf*E2bc was stable following elution in the presence of imidazole, however, following buffer exchange to remove imidazole, the protein aggregated. Regardless, the concentrations obtained (0.3 – 0.5 mg/ml) were ultimately sufficient for biochemical and structural analyses of the protein.

3.1.3.4.3 Purification of GST-r*Pf*E2c

Protein lysate from 2 L of GST-r*Pf*E2c expressed overnight at 18°C in BLR(DE3) was purified on a GSTrap FF column (GE Healthcare). Elution was carried out with 10 mM reduced glutathione and GST-r*Pf*E2c eluted in one peak (Fig. 3-17A & B).

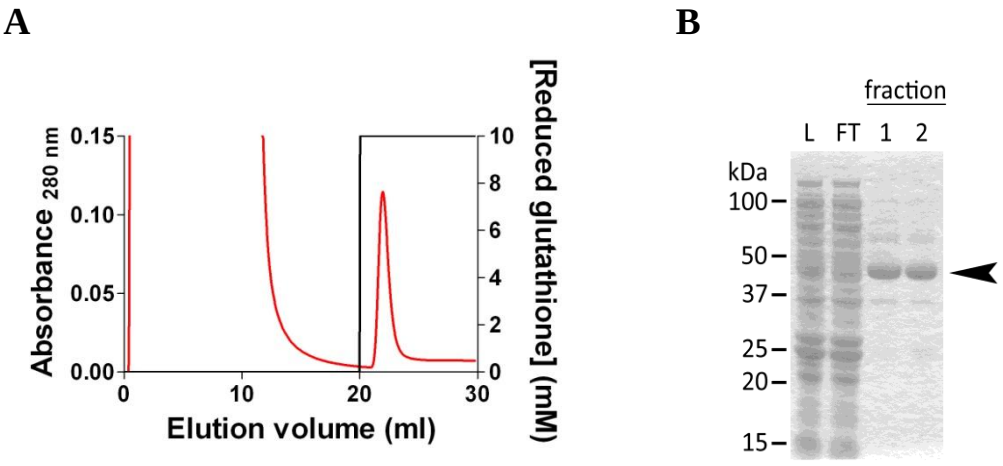


Figure 3-17. GSTrap purification of GST-r*Pf*E2c
(A) GST-r*Pf*E2c eluted in one peak (consisting of two fractions) from a GSTrap FF column with 10 mM reduced glutathione. (B) SDS-PAGE showed the presence of a 45 kDa protein (arrow) presumed to GST-r*Pf*E2c in both fractions (numbered). L, protein lysate; FT, flow through.

Cleavage of the GST-tag was attempted using PreScission Protease (GE Healthcare) (section 2.6.8.5). Cleavage was first tried on eluted GST-r*Pf*E2c with overnight incubation of the protein with 60 U of PreScission Protease at 4ºC. Cleavage of the recombinant protein from the GST-tag was successful (Fig. 3-18A), however, separation of the cleaved r*Pf*E2c was not (Fig. 3-18B & C). r*Pf*E2c did not elute in the flow-through as it should but a large fraction of it was still bound onto the GSTrap FF column or interacted with the bound GST-tag. r*Pf*E2c only eluted along with the GST-tag on addition of 10 mM reduced glutathione. On-column cleavage under the same conditions as above did not solve the problem. The cleaved r*Pf*E2c and GST-tag could not be separated (Fig. 3-18D & E) thus, further studies with this construct were not continued.

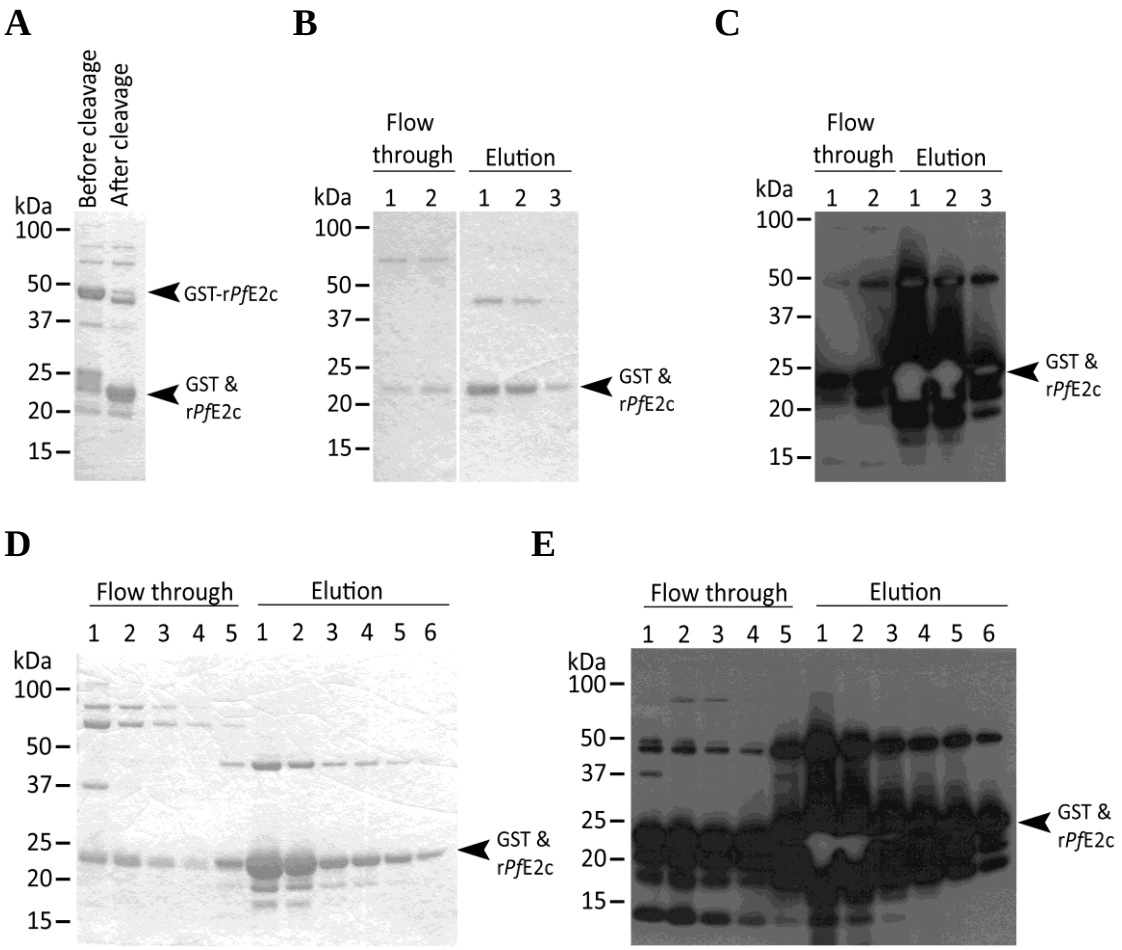


Figure 3-18. Cleavage of GST-tag from GST-r*Pf*E2c
(A) Before cleavage with PreScission Protease, a large amount of a 45 kDa protein corresponding to the size of GST-r*Pf*E2c can be seen. Following cleavage, only a small amount is left at 45 kDa as most of the protein has a mass of ~23 kDa. This size corresponds to both the cleaved r*Pf*E2c and GST-tag. (B) SDS-PAGE of flow-though and elution fractions (numbered) following attempted purification of r*Pf*E2c from the GST-tag. Very little 23 kDa protein is seen in the flow-though fractions. The elution fractions contain most of the protein. (C) Anti-GST-antibody western blot showing the presence of GST-tag in all fractions, thus confirming that r*Pf*E2c and the GST-tag have not been well separated. The anti-GST-antibody is extremely sensitive, even at a dilution of 1:50,000. (D) SDS-PAGE of on-column cleavage fractions (numbered). Again most of the 23 kDa protein is in the elution fractions. (E) An anti-GST-antibody western blot confirmed all fractions to still contain the GST-tag.

3.1.3.5 Kinetic analyses

During transfer of the acetyl group from E1 to CoA to form acetyl-CoA an intermediate, 8-acetyl-dihydrolipoamide, is formed which absorbs light at 240 nm. In the E2 activity assay, this reaction is reversed and coupled to the production of acetyl-CoA by phosphotransacetylase (PTA) (Fig. 3-19). Dihydrolipoic acic (DHLA) is added to reverse the reaction, as the lipoamide moiety on purified E2 would be in the oxidised form as required for the natural forward direction of the reaction. The increase in absorbance due to increasing quantities of the product 8-acetyl-dihydrolipoamide was measured at 240 nm. Activity is represented as ΔAbs_{240nm}/min/mg as the extinction coefficient for 8-acetyl-dihydrolipoamide cannot be determined precisely. This is because acetyl groups shuffle between sulphurs in acetylated dihydrolipoamide leading to a mixture of 8- and 6-acetyl dihydrolipoamide at a 3:1 ratio and also high levels of acetylation result in large amounts of a biacetyl derivative (O'Connor *et al.*, 1982; Yang and Frey, 1986; Yang *et al.*, 1997).

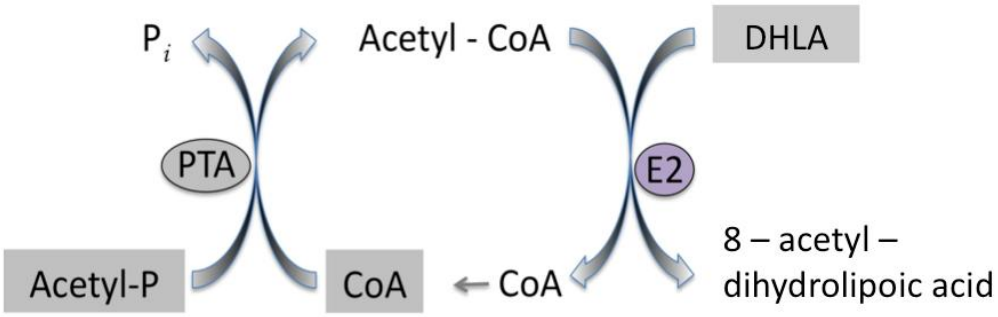


Figure 3-19. The E2 activity assay
PTA catalyses the formation of acetyl-CoA + P_i from acetyl-phosphate (Acetyl-P) and CoA. Upon addition of DHLA followed by E2, 8-acetyl-dihydrolipoamide + CoA is formed. CoA feeds back into the PTA reaction creating a cycle where acetyl-CoA is constantly produced.

3.1.3.5.1 Activity of His-r*Pf*E2m

Before each set of activity measurements, the appropriate controls to test each substrate and enzyme were carried out to ensure the reaction was working (Fig. 3-20). Special care was taken with DHLA, which was tested between measurements as well, because it started to oxidise once exposed to air. When using DHLA stored at 4°C overnight, an unstable

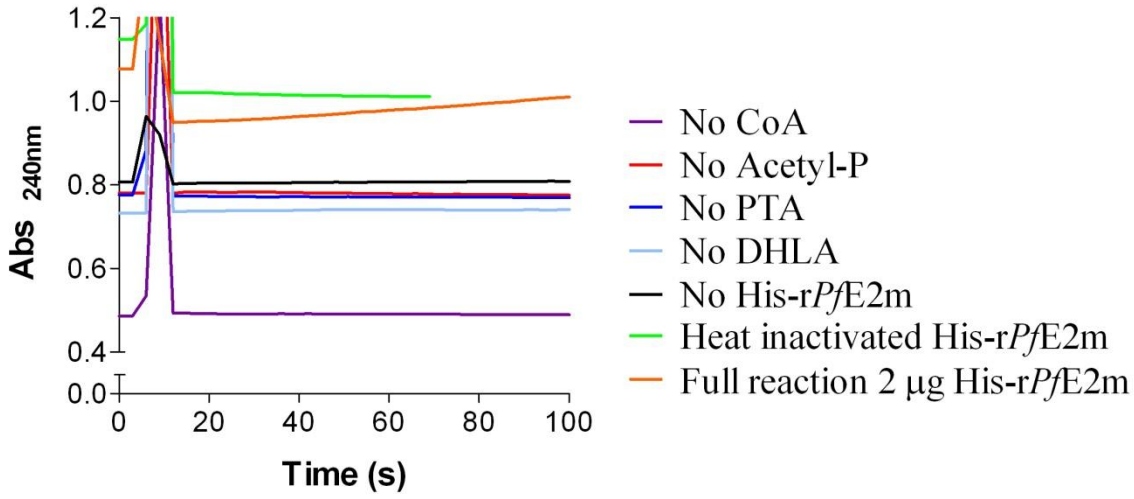


Figure 3-20. Control reactions for the E2 activity assays
To test the E2 activity assay, each component of the reaction was left out in turn and also heat inactivated (10 min at 100°C) His-r*Pf*E2m was tested. No increase in absorbance with time can be seen until the reaction contains all components and His-r*Pf*E2m.

absorbance prior to addition of His-r*Pf*E2m was observed. Thus, once a vial of DHLA was opened it was used for up to 8 h only to ensure the reaction was not compromised.

The activity values for His-r*Pf*E2m and His-r*Pf*E2bc were obtained by determining the steady state kinetics following addition of the recombinant protein into the reaction mixture. The rate was measured from the linear curve at a position where the change in absorbance over time did not vary significantly.

Activity of His-r*Pf*E2m was first assayed by increasing the amount of protein while both CoA and DHLA were kept at saturation levels (50 µM CoA, 1 mM DHLA). His-r*Pf*E2m activity increased in a linear fashion when 0.2 - 1 µg of protein was used (Fig. 3-21A). The pH optimum of His-r*Pf*E2m was determined to be between pH 7 and 8 (Fig. 3-21B) and there seemed to be a threshold of 0.75 mM for the amount of DHLA required for activity (Fig. 3-21C).

The activities of each of the His-r*Pf*E2m forms purified by gel filtration (Fig. 3-15C) were comparable, with peak 1 being slightly less active than the other three (Table 3-3). Kinetic parameters (Fig. 3-21D and Table 3-4) were determined for peak 3, as this contained the most protein and was used for subsequent structural characterisation.

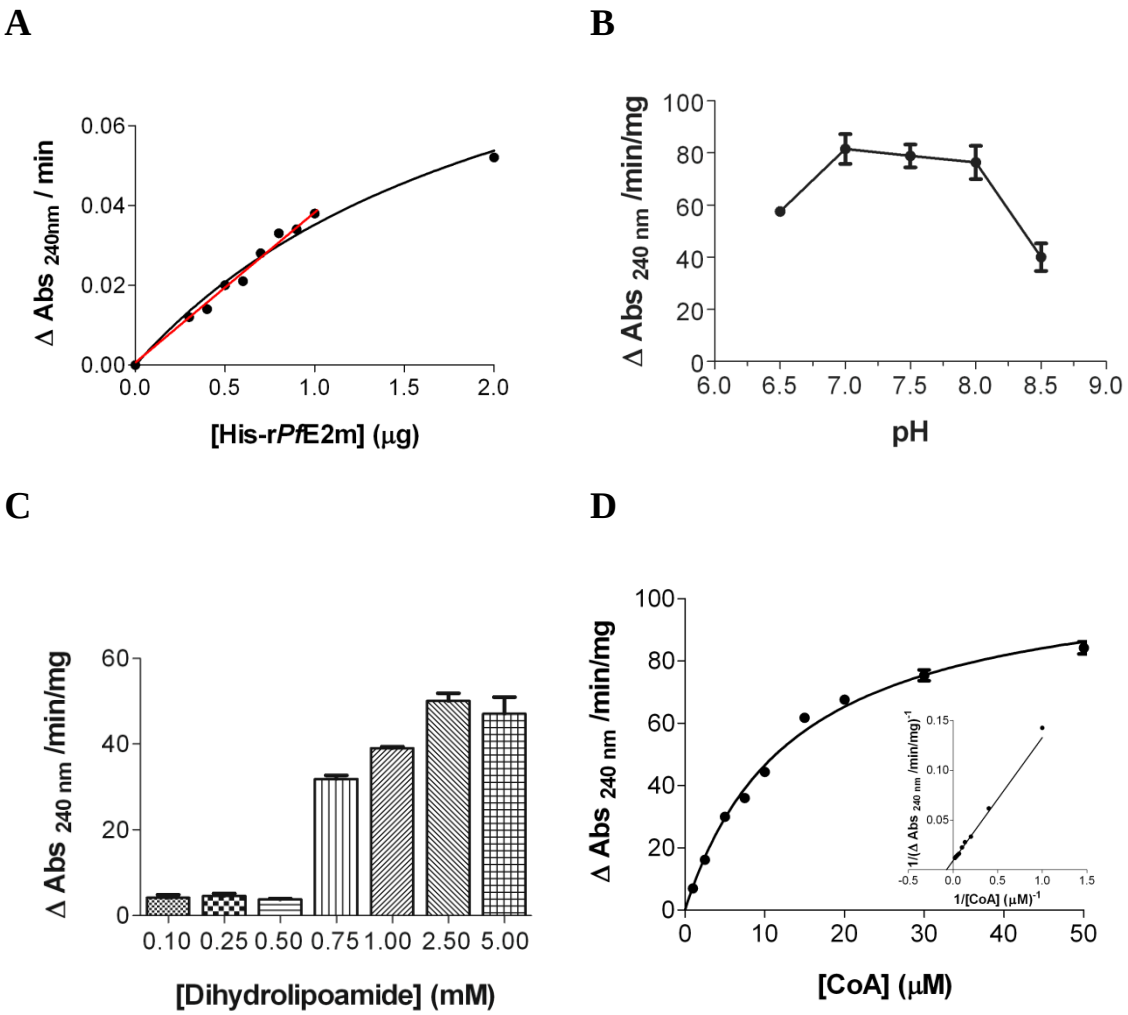


Figure 3-21. Activity of His-r*Pf*E2m

Each point represents the mean ± standard error of three independent measurements (unless otherwise stated). A total of 1 – 2 μg of His-r*Pf*E2m was used for each assay. (A) Activity of His-r*Pf*E2m increased as the amounts of protein were increased (0.2 – 2 μg). Only one measurement for each concentration was carried out for this data set. Both a linear (red) and non-linear (black) regression have been applied to the graph to show the linearity of the reaction up to 1 μg of protein. (B) The optimum pH of His-r*Pf*E2m was between pH 7 and 8. (C) The concentration of DHLA required in the reaction was determined by keeping CoA at 10 μM and varying the DHLA concentration (0.1 – 5 mM). (D) Michaelis-Menten kinetics of the His-r*Pf*E2m reaction for CoA. DHLA was kept constant at 1 mM while the concentration of CoA was varied (2.5 – 50 μM). The K_m was 13.5 ± 0.9 μM CoA and the v_{max} was determined as 110 ± 3 ΔAbs_{240nm}/min/mg by fitting the Michaelis-Menten and Lineweaver Burk equations in GraphPad Prism 5 (GraphPad Software Inc.). The inset shows the double reciprocal Lineweaver-Burk plot of the data.

3.1.3.5.2 Activity of His-r*Pf*E2bc

Both forms of His-r*Pf*E2bc purified by gel filtration (Fig. 3-16D) were active. Protein in peak 1 was slightly more active than in peak 2 (Table 3-3). The kinetic parameters of the two forms of His-r*Pf*E2bc for the substrate CoA were determined by keeping the concentration of DHLA constant (1 mM) and varying the concentration of CoA (2.5 – 50 μM) (Fig. 3-22). The values can be seen in Table 3-4.

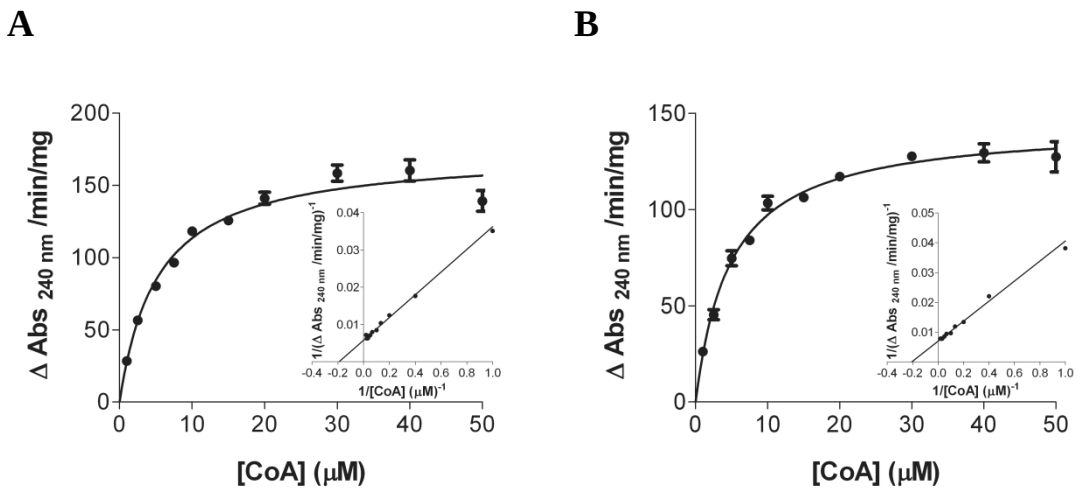


Figure 3-22. Activity of His-r*Pf*E2bc
Each point represents the mean ± standard error of three independent measurements. A total of 2 μg of His-r*Pf*E2bc was used for each assay. (A) The Michaelis-Menten plot for His-r*Pf*E2bc gel filtration peak 1. The kinetic parameters for CoA were: K_m = 5.3 ± 0.6 μM and v_{max} = 173 ± 5 ΔAbs $_{240nm}$/min/mg. (B) The Michaelis-Menten plot for His-r*Pf*E2bc gel filtration peak 2. The kinetic parameters for CoA were: K_m = 4.8 ± 0.4 μM and v_{max} = 144 ± 3 ΔAbs $_{240nm}$/min/mg. The insets in A and B show the double reciprocal Lineweaver-Burk plot of the data. Kinetic parameters were determined by fitting the Michaelis-Menten and Lineweaver Burk equations in GraphPad Prism 5 (GraphPad Software Inc.).

E2 source	ΔAbs_{240mn}/min/mg
P. falciparum, r*Pf*E2m-His	
Gel filtration peak 1	20 ± 0.8
peak 2	33 ± 0.4
peak 3	30 ± 0.9
peak 4	34 ± 0.4
P. falciparum, r*Pf*E2m-His	
Gel filtration peak 1	80 ± 4
peak 2	75 ± 2
P. falciparum, catalytic domain[1]	153.8 ± 16.2
Human E2[2]	19.4 ± 0.5
Bovine E2-E3BP[2]	12.3 ± 0.3
Maize mitochondrial PDC E2[3]	191
B. stearothermophilus[4]	7
A. vinelandii mature -length[5]	79.4
A. vinelandii catalytic domain[6]	117.9
E. coli catalytic domain[6]	47.1

Table 3-3. Comparison of E2 activities
Activities of the various His-r*Pf*E2m and His-r*Pf*E2bc species purified by gel filtration (Fig. 3-15C and 3-16D, respectively). Values are the mean ± standard error of three independent measurements. For comparison the activities of E2 from other sources are added. [1]Foth *et al*. (2005), [2]Yang *et al*. (1997), [3]Thelen *et al*. (1999), [4]Lessard *et al*. (1998), [5] Hanemaaijer *et al*. (1989), [6]Schulze *et al*. (1991b).

E2 source	v_{max} (ΔAbs_{240mn}/min/mg)	K_m (μM CoA)
P. falciparum, His-r*Pf*E2m	110 ± 3	13.5 ± 0.9
P. falciparum, His-r*Pf*E2bc		
Gel filtration peak 1	173 ± 5	5.3 ± 0.6
peak 2	144 ± 3	4.8 ± 0.4

Table 3-4. Comparison of kinetic parameters for CoA
The kinetic parameters for each of the measured His-r*Pf*E2m and His-r*Pf*E2bc samples. Values are the mean ± standard error of three independent measurements.

3.2 Dihydrolipoamide dehydrogenase

3.2.1 Introduction

Dihydrolipoamide dehydrogenase (E3) (EC 1.8.1.4) is part of the oxidoreductase family of proteins, which act on a sulphur group of donors and use NAD^+ or $NADP^+$ as an electron acceptor (EC 1.8.1.-). E3 catalyses the final step of the PDC reaction; the oxidation of the reduced lipoyl moiety of E2 using NAD^+ as the final electron acceptor (Fig. 3-23). For this reaction, E3 requires the cofactor flavin adenine dinucleotide (FAD) and redox disulphides to transfer the reducing equivalents from E2-bound dihydrolipoamide to NAD^+ (Patel *et al.*, 1995). The protein is a highly conserved homodimer (Mattevi *et al.*, 1992b).

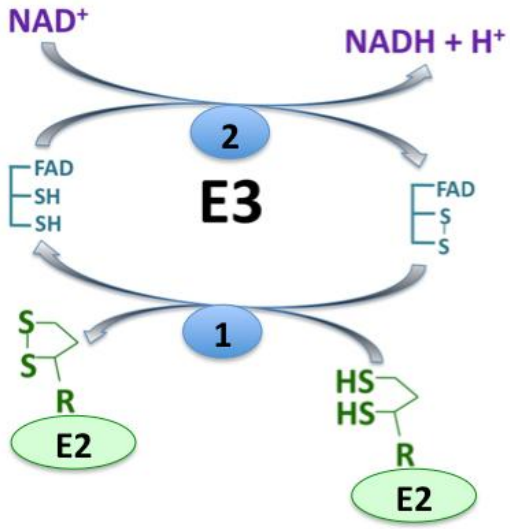


Figure 3-23. Reaction catalysed by E3
The reaction is a ping-pong mechanism characterised by two reactions. (1) E3 oxidises the E2 dihydrolipoamide moiety while reducing the disulphides in E3. (2) E3 is oxidised by transferring an electrons to NAD^+ and releasing a proton.

The gene for apicoplast *P. falciparum* E3 (*Pf*aE3) (PlasmoDB annotation PF3D7_0815900, www.plasmodb.org) is located on chromosome 8, is 2001 bp and encodes a 667 amino acid protein with a predicted molecular mass of 75.6 kDa (McMillan, 2006). *P. falciparum* has two distinct E3s. One localises to the apicoplast and is used by the PDC, while the

other is found in the mitochondrion and is shared by the other α-ketoacid dehydrogenase complexes (KADHs) (McMillan *et al.*, 2005). A similar situation is found in plants, in which the mitochondrion and the chloroplast have separate E3 proteins (Mooney *et al.*, 2002). The apicoplast *Pf*aE3 showed the highest amino acid sequence identity with plant and cyanobacterial E3 (37 – 38%) E3 (McMillan, 2006).

3.2.2 Aims

The aims for this part of the project were to:

- express recombinant *P. falciparum* E3 (r*Pf*aE3) in soluble form
- purify r*Pf*aE3
- determine the kinetic parameters of r*Pf*aE3

3.2.3 Results

3.2.3.1 Generation of expression constructs

As for *Pf*E2 the native coding sequence for *Pf*aE3 contained several rare codons and in addition, Shine-Dalgarno-like sequences (GGAGGA) at 406 – 411 bp and 484 – 489 bp. Thus codon optimisation of the full-length *Pf*aE3 was carried out by GENEART AG (from 2010 onwards the company is known as GeneArt® Gene Synthesis, Life Technologies). Following gene synthesis, the company cloned the mature-length *Pf*aE3 sequence (starting at Lys101) into the expression vector pET-15b, with a cleavable N-terminal 6 x His-tag producing the construct His-r*Pf*aE3-pET-15b. The starting residue of the mature-length *Pf*aE3, had been previously estimated by McMillan (2006). A construct with the first 120 amino acids of *Pf*aE3 fused to a GFP-tag with a predicted molecular mass of 42 kDa was transfected into *P. falciparum*. As the fusion protein was expressed in the parasite and transported into the apicoplast, the bipartite signal/transit peptide was cleaved. Immunoprecipitation was then carried out to isolate the fusion protein from the parasites and a 29 kDa GFP-tagged protein was observed. From this the deduction was made that the mature-length r*Pf*aE3 started around Met110 (McMillan, 2006). However, as this was only an estimation of the start site, to make sure none of the mature-length protein residues were missed, the His-r*Pf*aE3-pET-15b expression construct was started at Lys101.

3.2.3.2 Recombinant His-r*Pf*aE3 expression

An attempt to express the codon optimised His-r*Pf*aE3-pET-15b in Rosetta™ 2(DE3) *E. coli* at 30°C resulted in very small amounts of the 64 kDa protein in soluble form, which degraded 4 h following induction (Fig. 3-24A & B). Expression of the construct in BLR(DE3) or BL21 Star™(DE3) *E. coli* produced no soluble protein. Purification of His-r*Pf*aE3-pET-15b following expression at 37°C for 2 h, was unsuccessful as the protein did not bind the nickel affinity chromatography resin. The reason for the very low expression of soluble protein could have been the additional amino acids present at the beginning of the protein (K101-M110). These were likely to have been still part of the hydrophobic targeting peptide. Thus, a previously made construct (McMillan, 2006) containing the

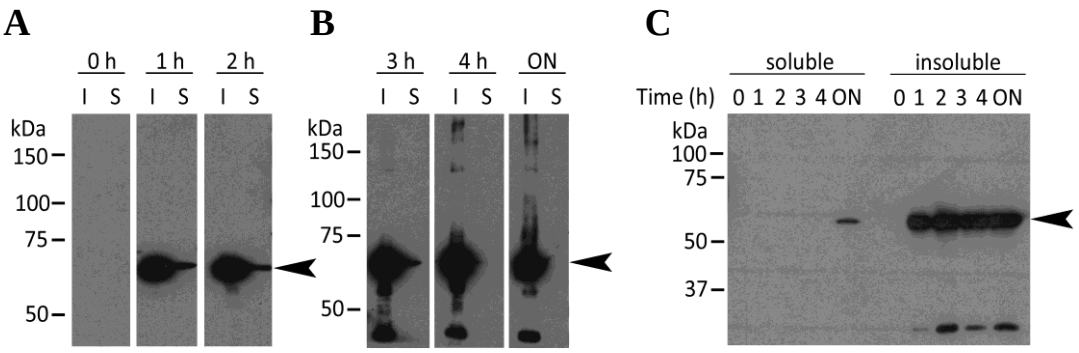


Figure 3-24. Western blot analysis of expression of r*Pf*aE3
Expression of His-r*Pf*aE3-pET-15b using Rosetta™ 2(DE3) *E. coli* was tested at 30°C following induction with 1 mM IPTG. Samples of expression cultures were taken (A) before induction at 0 h, after induction at 1 h, 2 h and (B) after 3 h, 4 h and overnight expression (ON). Both insoluble (I) and soluble (S) fractions were analysed with anti-His-tag antibody western blotting. A ~ 60 kDa protein (arrow) corresponding to His-r*Pf*aE3-pET-15b could be seen. C) Expression of His-r*Pf*aE3 (in pQE30) using NovaBlue(DE3) *E. coli* at 30°C as previously described by McMillan (2006). Anti-His-tag antibody western blotting showed a small amount of soluble degradation-free His-r*Pf*aE3, with a mass of ~ 60 kDa, after overnight expression.

native amino acid sequence of the mature-length r*Pf*aE3, starting at Met110, in the vector pQE30 was used (construct name His-r*Pf*aE3). An expression trial at 30°C in NovaBlue(DE3) *E. coli* showed expression only after overnight incubation (Fig. 3-24C), however, this was free of degradation products and the yield was greatly improved following large scale expression.

3.2.3.3 Recombinant His-r*Pf*aE3 purification

Protein lysate from 1 L of His-r*Pf*aE3-pQE30 culture expressed using NovaBlue(DE3) *E. coli* at 30°C overnight was incubated with 1 ml of Ni-NTA agarose. The mixture was poured into an empty column and the flow-through collected. The resin was washed two times with 20 mM imidazole, once with 50 mM imidazole and elution carried out with 250 mM imidazole. His-r*Pf*aE3 eluted mainly in the first elution fraction along with some degradation products (Fig. 3-25A & B).

Fractions from two separate purifications were pooled and loaded onto a HiLoad™ 16/60 Superdex 200 gel filtration column. His-r*Pf*aE3 eluted in one peak corresponding to molecular mass of 140 kDa (Fig. 3-25C & D). The protein concentration in the peak was 0.2 mg/ml. Concentration of the protein was carried out using aquacide (section 2.6.10).

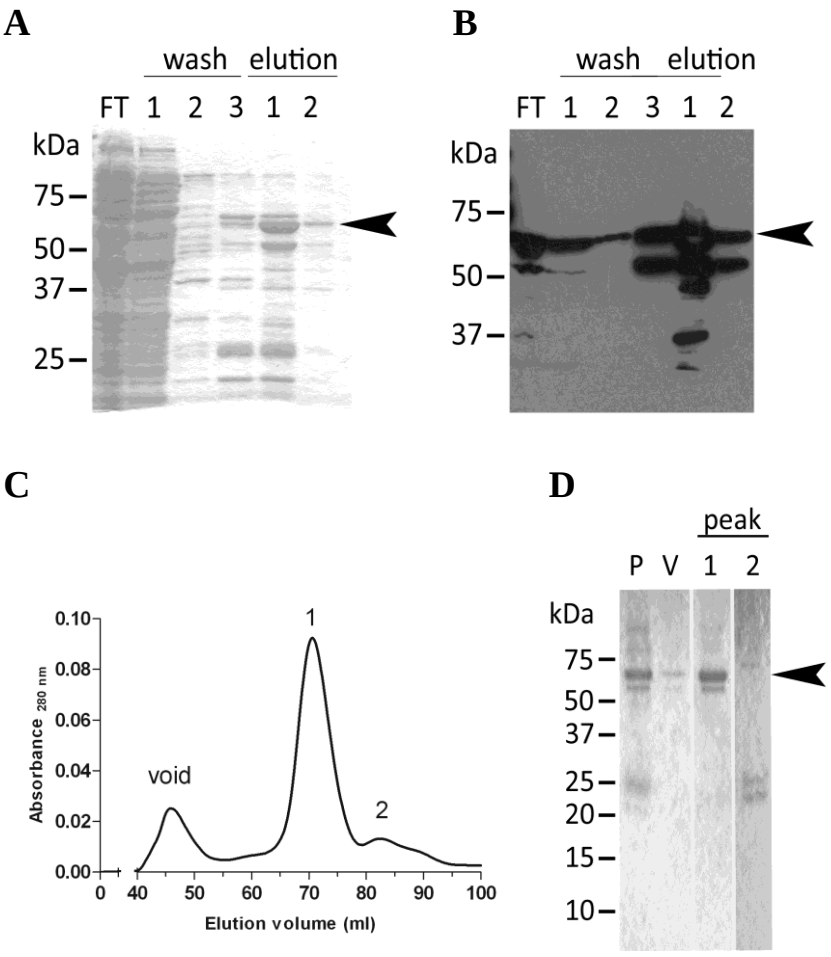


Figure 3-25. Ni-NTA batch purification of His-r*Pf*aE3
(A) SDS-PAGE of the flow-through (FT), wash and elution fractions. Numbers above the lanes denote fractions. In elution fraction 1, a ~ 60 kDa protein could be seen (arrow), which was confirmed to be His-r*Pf*aE3 by anti-His-tag antibody western blot (B). (C) The first elution fractions of two Ni-NTA purifications were loaded onto a HiLoad™ 16/60 Superdex 200 column. His-r*Pf*aE3 eluted in one peak corresponding to 140 kDa. For standard curve see Figure 3-13E. (D) SDS-PAGE shows that peak 1 contained the purified His-*rPf*aE3. FT, flow through; P, pooled Ni-NTA elutions; V, void volume. The SDS-PAGE and western blot in figures A & B were done by Jamie Whitelaw.

3.2.3.4 Kinetic analyses

The kinetic parameters of His-r*Pf*aE3 forward and reverse reactions were assayed at 25°C using 1 μg of protein as described in section 2.6.14. The change in absorbance at 340 nm due to reduction of NAD^+ or oxidation of NADH, was monitored at 240 nm for the forward and reverse reactions, respectively. The K_m and v_{max} for dihydrolipoamide (DHLA) were determined by keeping NAD^+ constant at 2 mM and varying the concentration of DHLA

between 50 µl and 2 µl mM (Fig. 3-26A). For the kinetic parameters for NAD^+, DHLA was kept constant at 2 mM and NAD^+ varied between 62.5 and 2 mM (Fig. 3-26B).

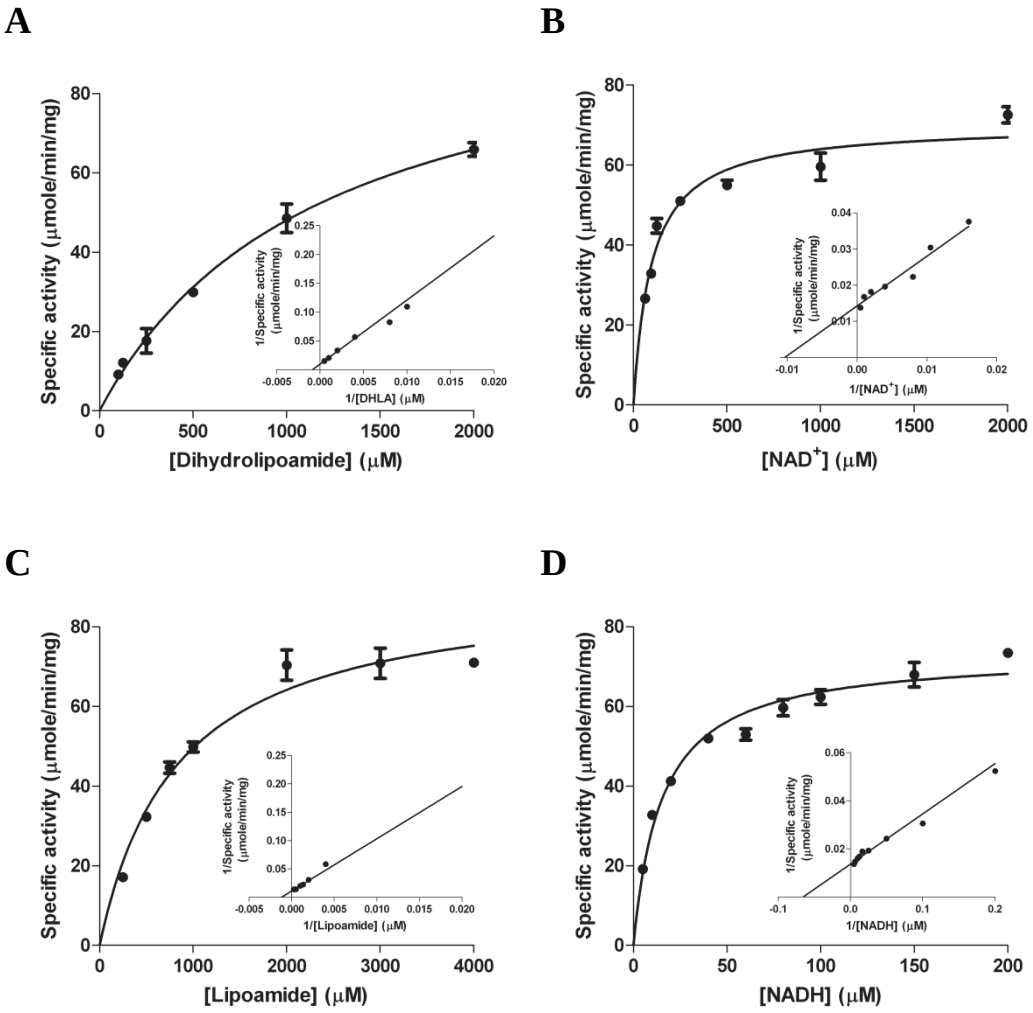


Figure 3-26. Activity of His-r*Pf*aE3
Each point represents the mean ± standard error of three independent measurements. The figures show the Michaelis-Menten plot for each substrate and the double reciprocal Lineweaver-Burk plot of the data as an inset. The K_m and v_{max} for each substrate was determined by fitting the Michaelis-Menten and Lineweaver Burk equations in GraphPad Prism 5 (GraphPad Software Inc.). (A) DHLA: K_m = 1160 ± 105 µM and v_{max} = 104 ± 5 µmol/min/mg (B) NAD^+: K_m = 96 ± 12 µM and v_{max} = 104 ± 5 µmol/min/mg (C) Lipoamide: K_m = 841 ± 97 µM and v_{max} 91 ± 4 µmol/min/mg (D) NADH: K_m = 15 ± 2 µM and v_{max} = 73 ± 2 µmol/min/mg.

To determine the K_m and v_{max} for lipoamide the concentration of NADH was kept constant at 200 µM and lipoamide was varied between 50 µM and 4 mM (Fig. 3-26C). Finally for NADH kinetic parameters, lipoamide was kept constant at 2 mM and NADH was varied

between 10 μM and 200 μM (Fig. 3-26D). The kinetic parameters were comparable to E3 from other species (Table 3-5).

	His-r*Pf*aE3	*P. falciparum* mitochondrial E3[1]	Human E3[2]	Trypanosome E3[3]
K_m DHLA	1160 ± 105 μM	146 ± 15 μM	570 μM	130 μM
K_m NAD^+	96 ± 12 μM	450 ± 30 μM	290 μM	600 μM
K_m lipoamide	841 ± 97 μM	870 ± 270 μM	1010 μM	800 μM
K_m NADH	15 ± 2 μM	21 ± 6 μM	51 μM	25 μM

Table 3-5. Comparison of the K_m of E3 from various species
His-r*Pf*aE3 values are the mean ± standard error of three independent measurements. [1]McMillan *et al*. (2005), [2]Kim *et al*. (1991), [3]Schöneck *et al*. (1997).

3.3 Summary

- Full-length *Pf*E2 had low sequence identity compared with PDC E2 from other organisms, especially at the N-terminus where there is great variation between the species due to different numbers of LDs. The linker regions also had low identity, for example, the human E2 linkers had high levels of alanine and proline whereas the *Pf*E2 linkers were abundant in glycine, lysine and asparagine.

- Regardless of the low amino acid sequence identity, *Pf*E2 domains contain several important conserved residues. The LDs contain the lysine required for lipoylation (Lys93 and Lys 223), the SBD contains an arginine (Arg383) required for E3 binding in other organisms, and the CD contains the conserved DHR motif with the histidine (His612) essential for activity. Some key residues in the *Pf*E2 domains differ from the human counterparts and could thus be of interest in future studies.

- The domains of various *Plasmodium* E2 are quite conserved, while differences are seen in the linkers. *Pf*E2 did not have more sequence identity with the other human *Plasmodium* parasites nor the rodent species but seemed to be a mix of the two groups.

- *E. coli* codon optimised mature-length *Pf*E2 (His-r*Pf*E2m) was expressed in Rosetta™ 2(DE3)pLysS *E. coli* at 18°C overnight to obtain a final yield of 1.8 mg/L of baterial culture. His-r*Pf*E2m was purified using $ZnSO_4$ POROS® PEEK™ column and gel filtration. The purified protein could be concentrated to 3.5 mg/ml. His-r*Pf*E2m was more active than human E2.

- *E. coli* codon optimised truncated *Pf*E2 comprising the SBD and CD (His-r*Pf*E2bc) was expressed in BLR(DE3) *E. coli* at 25°C overnight to obtain a final yield of 0.25 mg/L of bacterial culture. The recombinant protein was purified using a HisTrap FF column and gel filtration. Concentration of His-r*Pf*E2bc was difficult as the protein aggregated at above 0.7-0.8 mg/ml, thus His-r*Pf*E2bc was only concentrated to 0.5 mg/ml. His-r*Pf*E2bc was more active than His-r*Pf*E2m.

- An attempt was made to purify a GST-tagged *Pf*E2 CD (GST-r*Pf*E2c). Cleavage of the tag from the recombinant protein was successful, however, the GST-tag could not be separated from r*Pf*E2c. Thus no further work was done with this construct.

- Native mature-length apicoplast *Pf*E3 (His-r*Pf*aE3) was expressed in NovaBlue(DE3) at 30°C overnight. A final yield of 1.5 mg/L of bacterial culture was obtained. His-r*Pf*aE3 was purified using the Ni-NTA batch purification method and could be concentrated to 2.9 mg/ml. The recombinant protein activity was comparable to that of E3 from other organisms.

Chapter 4

Structural characterisation of recombinant *P. falciparum* dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase

4.1 Dihydrolipoamide acetyltransferase

4.1.1 Introduction

E2 has a multidomain structure; *Pf*E2 contains two lipoyl-domains (LD) at the N-terminus followed by an E1/E3 sub-unit binding domain (SBD), and a catalytic domain (CD) at the C- terminus, all separated by flexible linkers (Fig. 4-1). Three E2 monomers assemble into a tightly bound trimer. For example, in *A. vinelandii* trimer formation involves 66 residues per monomer and 48 hydrogen bonds (i.e. 16 hydrogen bonds per monomer) (Mattevi *et al.*, 1992a; Mattevi *et al.*, 1993a). The active sites are formed at the interface of each monomer. The interactions between the trimers to form the large E2 core are not as strong as the intra-trimer bonds. Core formation involves only a hydrophobic 3_{10}-helix "ball" at the C-terminus of the protein, which inserts into a hydrophobic "socket" and is kept in place by only two hydrogen bonds (Mattevi *et al.*, 1992a; Izard *et al.*, 1999). The E2 core is an icosahedral 60-mer in humans, Gram-positive bacteria and plant mitochondria, and an octahedral 24-mer in Gram-negative bacteria and the plant chloroplast (Mattevi *et al.*, 1992a; Izard *et al.*, 1999; Mooney *et al.*, 1999; Yu *et al.*, 2008b).

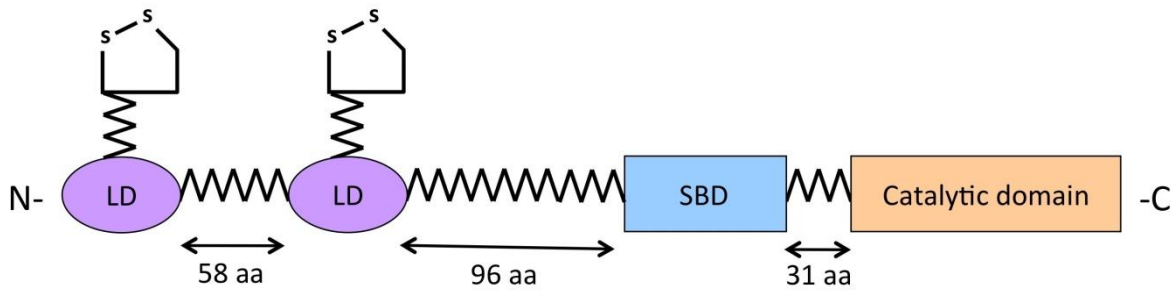


Figure 4-1. Multidomain structure of *Pf*E2
The linker between the two lipoyl-domains (LD) is 58 amino acids (aa), between the second LD and the E1/E3 sub-unit binding domain (SBD) 96 residues and between the SBD and catalytic domain (CD) 31 amino acids.

The structure and stochiometry of the *Plasmodium* E2 core are unknown. As the apicoplast is of red algal origin and a non-photosynthetic chloroplast-like organelle (Botte *et al.*, 2011), the hypothesis tested here was that *Pf*E2 forms a 24-mer. Full-length E2 has not yet been crystallised, probably due to the flexible linkers. Thus, to obtain structural information on *Pf*E2, other biophysical methods were utilised to obtain structural information. In this study, analytical ultracentrifugation (AUC) and small angle X-ray scattering (SAXS) were used to investigate the solution structure and stoichiometry of recombinant *Pf*E2.

4.1.2 Aims

The aims for this part of the project were to:

- determine the stoichiometry of the r*Pf*E2 core
- obtain a structural model of r*Pf*E2

4.1.3 Results

4.1.3.1 Homology modelling of *P. falciparum* E2 domains

In the absence of a crystal structure for *Pf*E2, homology modelling was carried out to obtain an indication of the *Pf*E2 structure. The I-TASSER server for protein structure and function prediction was used for this (http://zhanglab.ccmb.med.umich.edu/I-TASSER/, Zhang, 2008; Roy *et al.*, 2010, section 2.8.2). I-TASSER generates structural models using templates from proteins in the PDB database that have similar folds and secondary structures as those predicted from the submitted amino acid sequence (Zhang, 2008). Due to the flexible linkers between the domains, I-TASSER was unable to model the mature length protein (residues 51-640). Thus, the individual domains were modelled.

The top three templates used for structure prediction of *P. falciparum* LD1 (*Pf*LD1) were the LD from the human E3 binding protein (E3BP) (PDB entry 2DNC, no associated publication), LD2 from human E2 (PDB entry 1FYC, Howard *et al.*, 1998), and the *A. vinelandii* α-ketoglutarate dehydrogenase complex E2 LD (PDB entry 1GHK, Berg *et al.*, 1996). Only one model with a high C-score (1.22, on a scale of -5 to 2, where a higher score indicates a better model) was generated (Fig. 4-2A), suggesting *Pf*LD1 is structurally very similar to that of other organisms (Fig. 4-2C). Root mean square deviations (rmsd) values showed *Pf*LD1 to be most similar to human E3BP LD (0.9 Å), human E2 LD2 (1.0 Å), *A. vinelandii* LD1 (1.3 Å), *G. stearothermophilus* LD (1.5 Å) and finally *E. coli* LD3 (2.2 Å). For reference, two identical structures will have a rmsd of 0.

Using the same templates as above, four I-TASSER models were obtained for *Pf*LD2 with C-scores ranging from 0.52 to -3.26. The best model, i.e. the model with the highest C-score, is shown in Fig. 4-2B. The amino acid sequence of *Pf*LD2 is less conserved than that of *Pf*LD1, however, the best model was still similar to *Pf*LD1 (rmsd = 0.8 Å). *Pf*LD2 was most similar to the human E2 LD2 and *A. vinelandii* LD1 (rmsd for both = 1.1 Å). It was also more similar to the other bacterial structures than was *Pf*LD1 (*G. stearothermophilus* LD rmsd: 1.1 Å, *E. coli* LD3 rmsd: 1.5 Å).

In both the *Pf*LD1 and *Pf*LD2 models, the lysine residue (Lys93 and Lys223, numbering based on full-length *Pf*E2 sequence) required for lipoylation of E2 is present in the same

loop as in E2 lipoyl-domains from other organisms. In the model for *Pf*LD1, the side chain of Lys93 points into the domain itself, which would not be ideal for lipoylation. This is likely a limitation of the model, as the residues surrounding Lys93 are similar to those observed in E2 LDs from other species. Also, the amino acid sequence alignments by I-TASSER of both *P. falciparum* LDs with their template sequences were the same as earlier alignments done with ClustalW2 (Fig. 3-2), with the lysines positioned within the same conserved DKA/S motif.

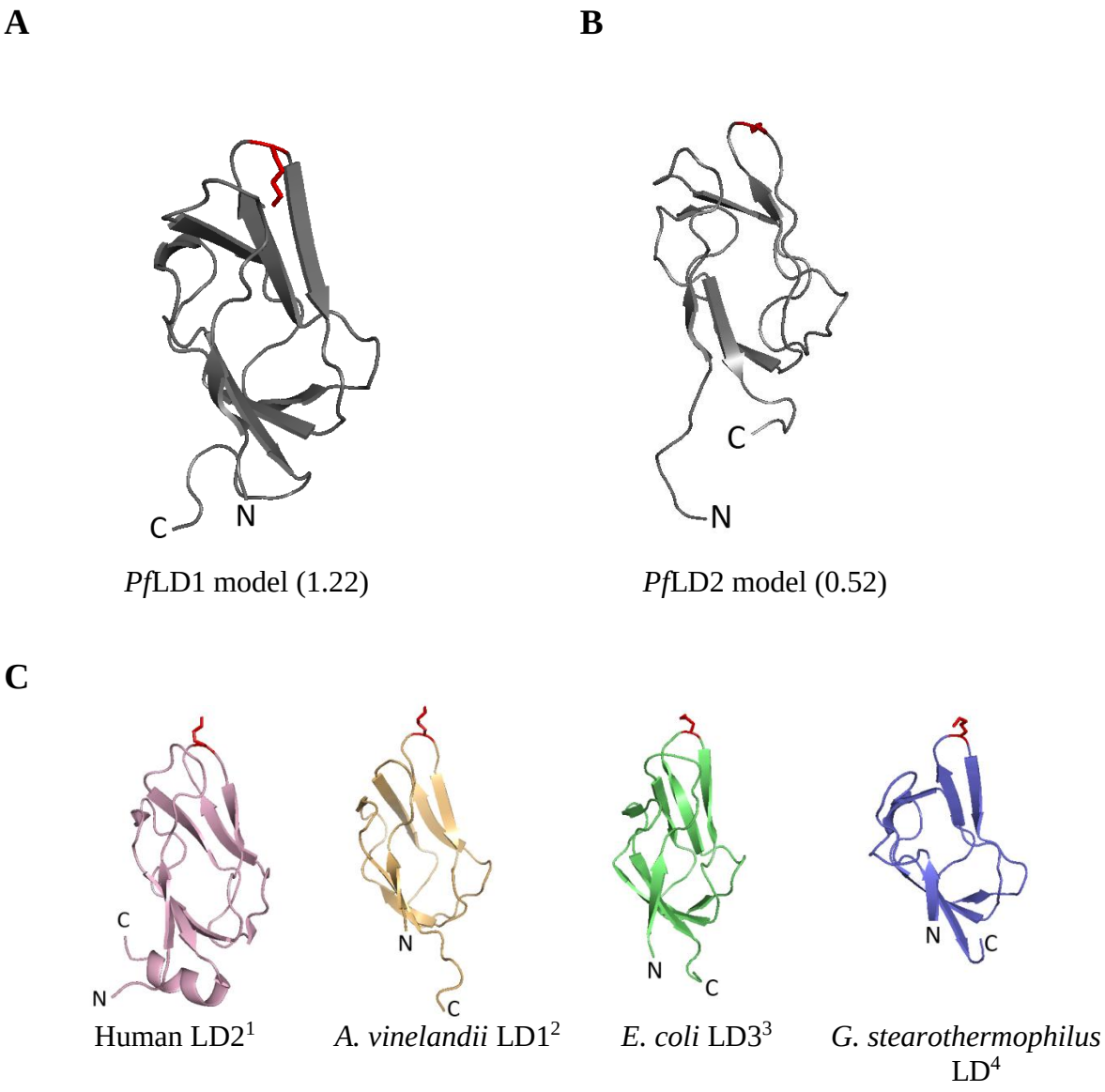
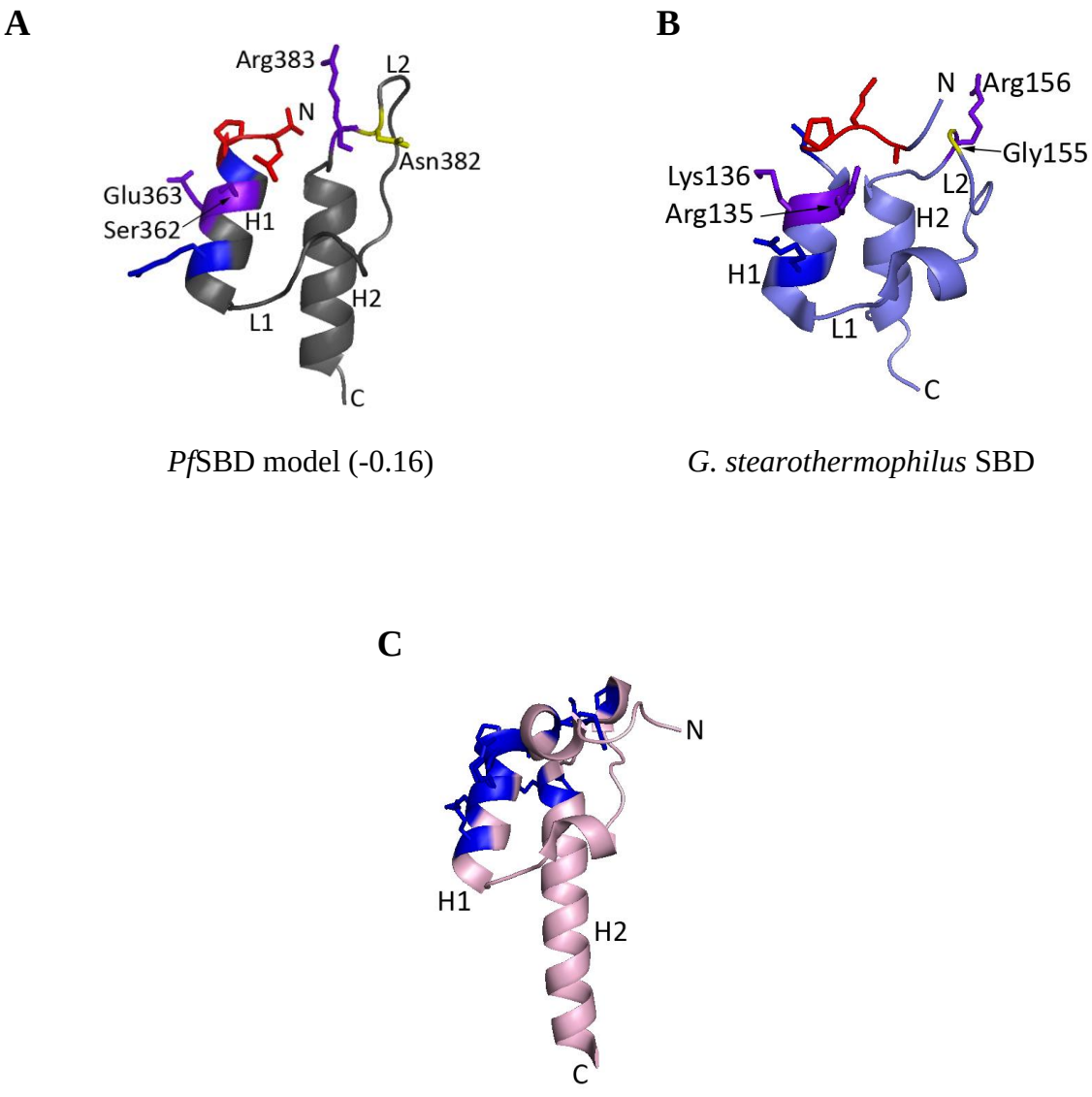


Figure 4-2. Models for *P. falciparum* lipoyl-domains
(A) I-TASSER generated only one model for *Pf*LD1. (B) The best model for *Pf*LD2. (C) NMR structures for lipoyl domains from other species. [1]PDB entry 1FYC, Howard *et al.* (1998); [2]PDB entry 1IYU, Berg *et al.* (1997); [3]PDB entry 1QJO, Jones *et al.* (2000); [4]PDB entry 1LAC, Dardel *et al.* (1993). C-scores of I-TASSER models are shown in parenthesis. Side chain of lysine residue required for lipoylation of E2 shown in red.

I-TASSER generated five models for the *P. falciparum* SBD (*Pf*SBD) with C-scores of -0.16 to -2.61. The best model is shown in Fig. 4-3A. The top three templates used were the SBD of *E. coli* succinyltransferase from the α-ketoglutarate dehydrogenase complex (PDB entry 2WXC, Neuweiler *et al.*, 2009), *Thermus thermophilus* E2 (PDB entry 2EQ9, no associated publication) and human branched-chain α-ketoacid dehydrogenase complex E2 (PDB entry 1ZWV, no associated publication). The resulting *P. falciparum* SBD model was, however, structurally most similar (rmsd=1.1 Å) to the human E3BP SBD (PDB entry 2F60, Brautigam *et al.*, 2006) and the SBD of the archaeum *Pyrobaculum aerophilum* E2 (PDB entry 1W4I, Ferguson *et al.*, 2005) (rmsd=1.2 Å). Alignment of the *Pf*SBD amino acid sequence with those of the templates was the same as alignments done with ClustalW2 (Fig. 3-4). Interesting observations can be made when mapping onto the *P. falciparum* I-TASSER model the residues reported to be involved in E1 and E3 binding in *G. stearothermophilus* (Mande *et al.*, 1996; Jung *et al.*, 2003; Frank *et al.*, 2005, Table 4-1). Arg383, a residue key in binding both E1 and E3, is in a similar position on loop L2 in the *P. falciparum* model as in *G. stearothermophilus* (Fig. 4-3A & B). Preceding the arginine in *G. stearothermophilus*, is a glycine (Gly155, in yellow, Fig. 4-3B), which is conserved in other organisms apart from *P. falciparum*, which has an asparagine instead (Asn382, in yellow, Fig. 4-3A). Kalia *et al.* (1993) suggested the glycine to be conserved in all organisms, as side-chains from any other amino acids could interfere with recognition of the arginine by E3. However, this does not seem to be the case in the homology model of *P. falciparum* SBD as the side-chain of Asn382 points in the opposite direction to that of Arg383. A striking difference is seen in the residues on helix H1 involved in both E1 and E3 binding. While *G. stearothermophilus* has Arg135 and Lys136, which have positively charged side-chains, *P. falciparum* has small, uncharged Ser362 and negatively charged Glu363 in these positions. This could suggest that a different set of residues are involved in the interaction between *P. falciparum* SBD and E1 and E3. Although the homology model was structurally similar to the SBD of the human E3BP (Fig. 4-3C), the residues involved in the interaction between the domain and E3 (Brautigam *et al.*, 2006; Ciszak *et al.*, 2006) are very different (Table 4-2). There is no crystal structure showing the interaction between E1 and the human E2 SBD, however, mutagenesis studies suggest residues similar to those in *G. stearothermophilus* are involved (Korotchkina and Patel, 2008). This is very interesting because if there is a sufficient difference in the binding mechanism of E1 and/or E3 in *P. falciparum* PDC

compared with the human proteins, this could be exploitable for development of drugs that target the *P. falciparum* PDC.



*Pf*SBD model (-0.16)

G. stearothermophilus SBD

Human E3BP SBD

Figure 4-3. Model of *P. falciparum* SBD
(A) The best I-TASSER model for *Pf*SBD, C-score is shown in parenthesis. (B) NMR structure for the *G. stearothermophilus* E1/E3BD, PDB entry 2PDD, Kalia *et al.* (1993). (C) Crystal structure of the SBD of human E3BP, PDB entry 2F60, Brautigam *et al.* (2006). Side chain residues in blue are those that interact with E3, in red with E1 and in purple those that interact with both E1 and E3. Residue in yellow is the highly conserved glycine in the SBD, which is an asparagine in *P. falciparum*. The residue positions in the *P. falciparum* I-TASSER model are based on the position of *G. stearothermophilus* residues reported to be involved in E1 and E3 binding.

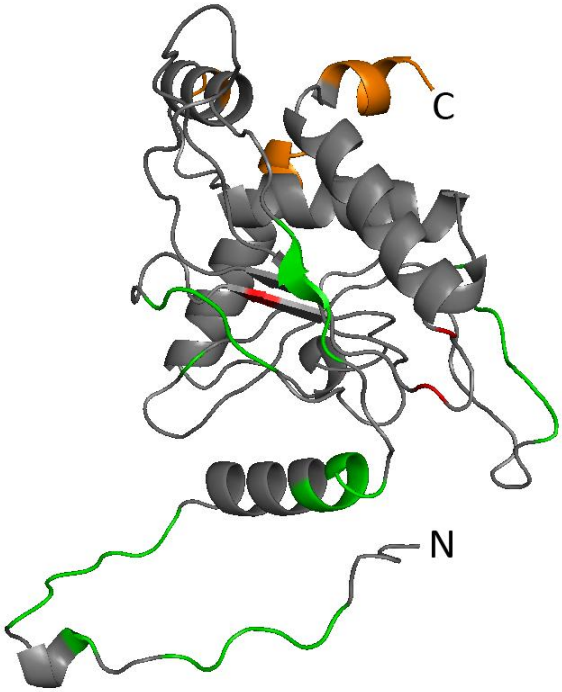
	G. stearothermophilus	*P. falciparum* model
E1 binding only	Ala130	Val357
	Met131	Leu358
	Pro132	**Pro359**
E3 binding only	Ser133	Ser360
	Arg139	**Arg366**
E1 and E3 binding	Arg135	Ser362
	Lys136	Glu363
	Arg156	**Arg383**

Table 4-1. SBD residues involved in binding E1 and E3
Shown are the *G. stearothermophilus* (Mande *et al.*, 1996; Jung *et al.*, 2003; Frank *et al.*, 2005) and the respective residues in *P. falciparum* based on sequence alignments (section 3.1.3.1) and structural position in homology model. Identical residues in bold.

	Human E3BP SBD	*Pf*E2 SBD model
E3 binding	Arg130	Met356
	Ser132	Leu358
	Pro133	**Pro359**
	Ala134	Ser360
	Arg136	Ser362
	Asn137	Glu363
	Glu140	Arg366
	Pro154	**Pro381**
	Ile157	**Ile384**
	Lys160	Glu387
	Glu161	Asp388

Table 4-2. Residues involved in human E3BP SBD interaction with E3
Human E3BP residues (Brautigam *et al.*, 2006; Ciszak *et al.*, 2006) and the respective residues in *P. falciparum* E2 SBD based on sequence alignments (section 3.1.3.1) and structural position in homology model. Apart from the charged amino acids, substitutions in the parasite SBD compared with the human E3BP SBD are not conserved. Identical residues in bold.

Four I-TASSER models were obtained for the *P. falciparum* CD (*Pf*CD), however, only one model had a good C-score (1.54) and was used for further analyses (Figure 4-4). The top templates used for the modelling were the *A. vinelandii* CD (PDB entry 1EAF, Mattevi *et al.*, 1993a) and human CD (PDB entry 3B8K, Yu *et al.*, 2008b). The resulting homology model was more similar to *G. stearothermophilus* (rmsd = 1.7 Å) than *A. vinelandii* (rmsd = 2.4 Å) and human CD (rmsd = 2.5 Å). Superimposition of the *P. falciparum* CD model onto these structures showed more overall similarity with bacterial than with human structures (Fig. 4-5).



*Pf*CD model (1.54)

Figure 4-4. I-TASSER model for *Pf*CD
C-score is shown in parenthesis. Positions of active site residues are in red, residues involved in trimer formation are in green and in orange, residues involved in n-mer core formation. The positions of these residues are based on *A. vinelandii* and *G. strearothermophilus* structures (Mattevi *et al.*, 1993a; Izard *et al.*, 1999).

A **B**

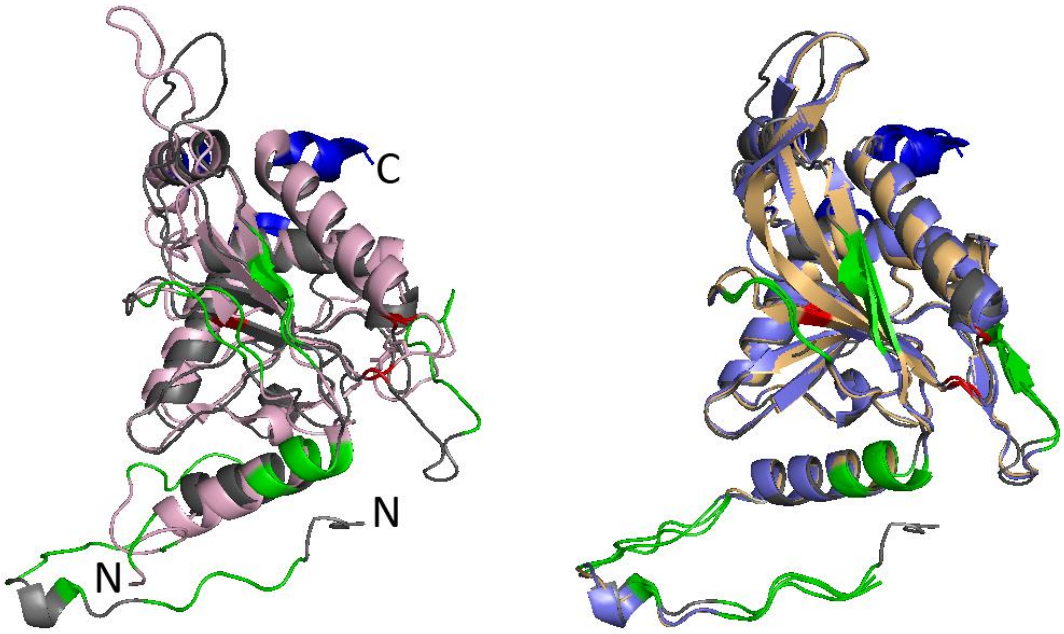


Figure 4-5. Alignment of *Pf*CD homology model (grey) with CD from other organisms
(A) Alignment with cryo-electron microscopy based structure for human CD (pink, PDB entry 3B8K, Yu *et al.*, 2008b). (B) Alignment with CD crystal structures for *A. vinelandii* (light orange) and *G. strearothermophilus* (light blue), PDB entry IEAE, Mattevi *et al.* (1993a) and 1B5S, Izard *et al.* (1999), respectively. Positions of active site residues are in red, residues involved in trimer formation in green and in dark blue, residues involved in n-mer core formation.

Based on the orientation of the *A. vinelandii* and *G. stearothermophilus* trimer crystal structures, the *Pf*CD homology structure was modelled into a trimer to assess how the residues thought to be involved in trimer formation are positioned (Mattevi *et al.*, 1993a; Izard *et al.*, 1999). Alignment of the *Pf*CD amino acid sequence with those of the templates was very similar to alignments done with ClustalW2 (Fig. 3-4). Conserved residues including the active site DHRXXD/N/YG motif as well as residues that could be involved in inter- and intratrimer bond formations were in the same positions in both alignments. As the overall structure of the *P. falciparum* homology model and bacterial CDs are very similar, the residues were mapped to analogous positions. There are four major points of interaction based on the *A. vinelandii* crystal structure. The residues in these areas are not conserved, however, they contain several polar, hydrophobic and charged amino acids, which could take part in interactions. On the outer surface of the trimer, in the middle, N-terminal residues 405-410 (numbering based on full-length *Pf*E2) of monomer A interact with residues 436-440 of monomer B helix H1. Towards the side, residues 418-424 of monomer A interact with residues 496-502 of monomer B (Fig. 4-6A). On the inner surface, which points towards the centre of the core complex, residues 594-600 form loops,

which reach over the neighbouring monomer surface (Fig 4-6B). Inside the monomer, residues 445-448 of monomer A interact with residues 568-572 of monomer C (Fig. 4-6C).

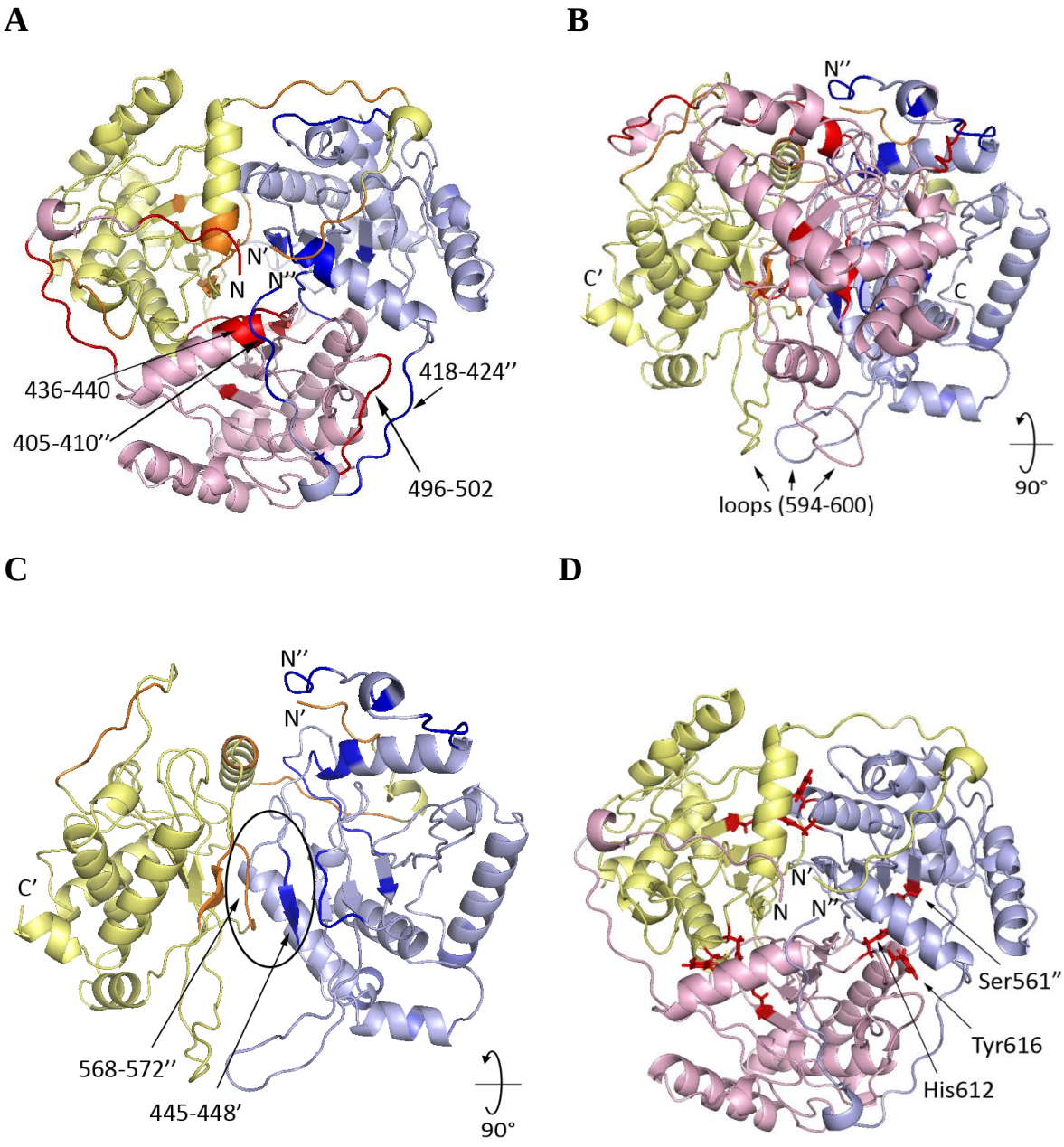


Figure 4-6. Model of *Pf*CD trimer
Each monomer is shown in a light colour (pink, light yellow, light blue). Residues involved in trimer interactions based those in *A. vinelandii* and *G. strearothermophilus* (Mattevi *et al.*, 1993a; Izard *et al.*, 1999) are in dark colours (red, orange, blue). (A) View of the trimer from above along the 3-fold axis. (B) Side view of model showing the loops reaching over the neighbouring monomer. (C) Inside view of the trimer showing interaction between two monomers (circled). (D) Active sites are formed at the interface of two monomers. Residues involved in catalytic activity are in red. Residue numbers shown are based on the full-length *P. falciparum* amino acid sequence.

The interactions in the human CD are similar apart from those in the N-terminal, as it is positioned differently (Yu *et al.*, 2008b). The trimer model also shows the formation of the active sites in the interface of two monomers (Fig. 4-6D) as in other organisms.

Finally, I analysed the potential interactions between the trimers to form the multimeric E2 core. Based on the residues determined for *A. vinelandii* and *G. strearothermophilus* (Mattevi *et al.*, 1992a; Izard *et al.*, 1999), *P. falciparum* residues 635-640 form the 3_{10}-helix "ball" at the C-terminus of one monomer in a trimer, could insert into the "socket" formed by residues Leu458, Val462, Ile529 and Tyr530 on the neighbouring trimer (Fig. 4-7A & B).

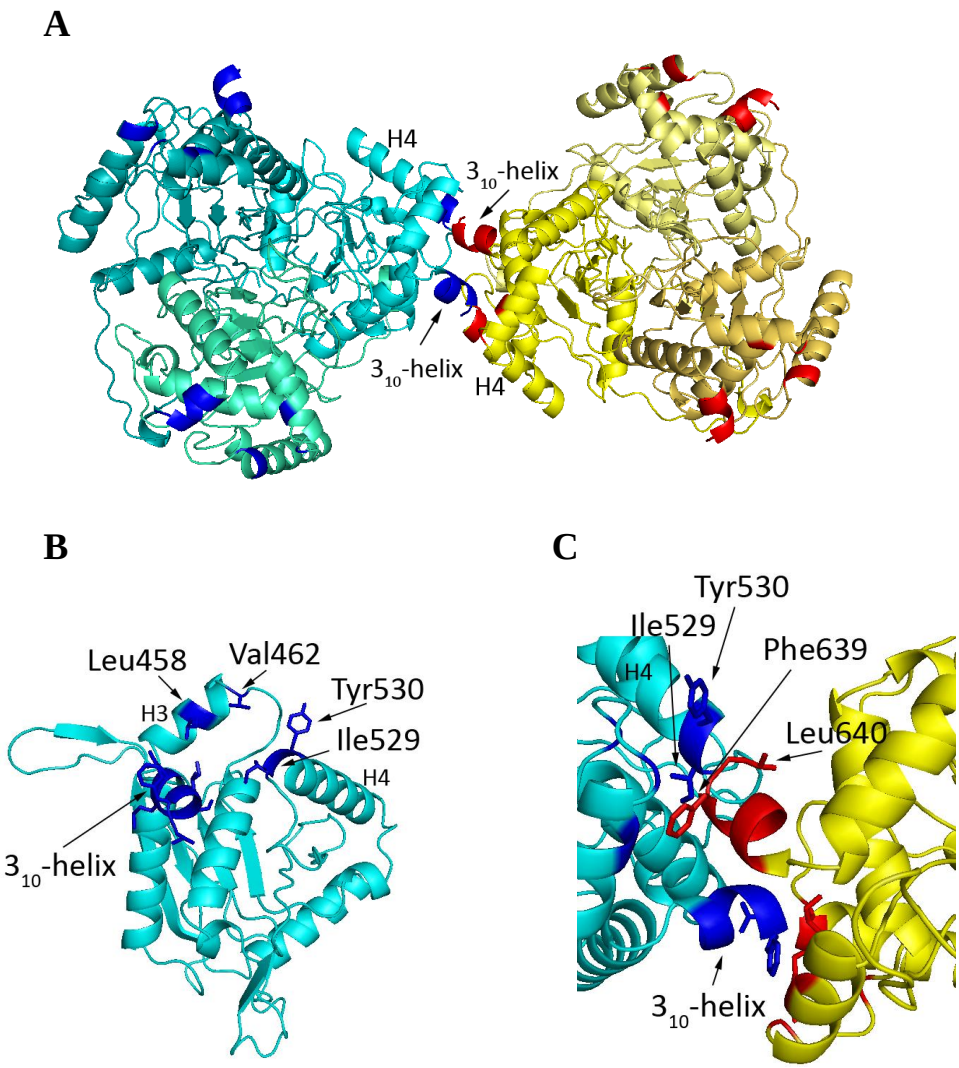
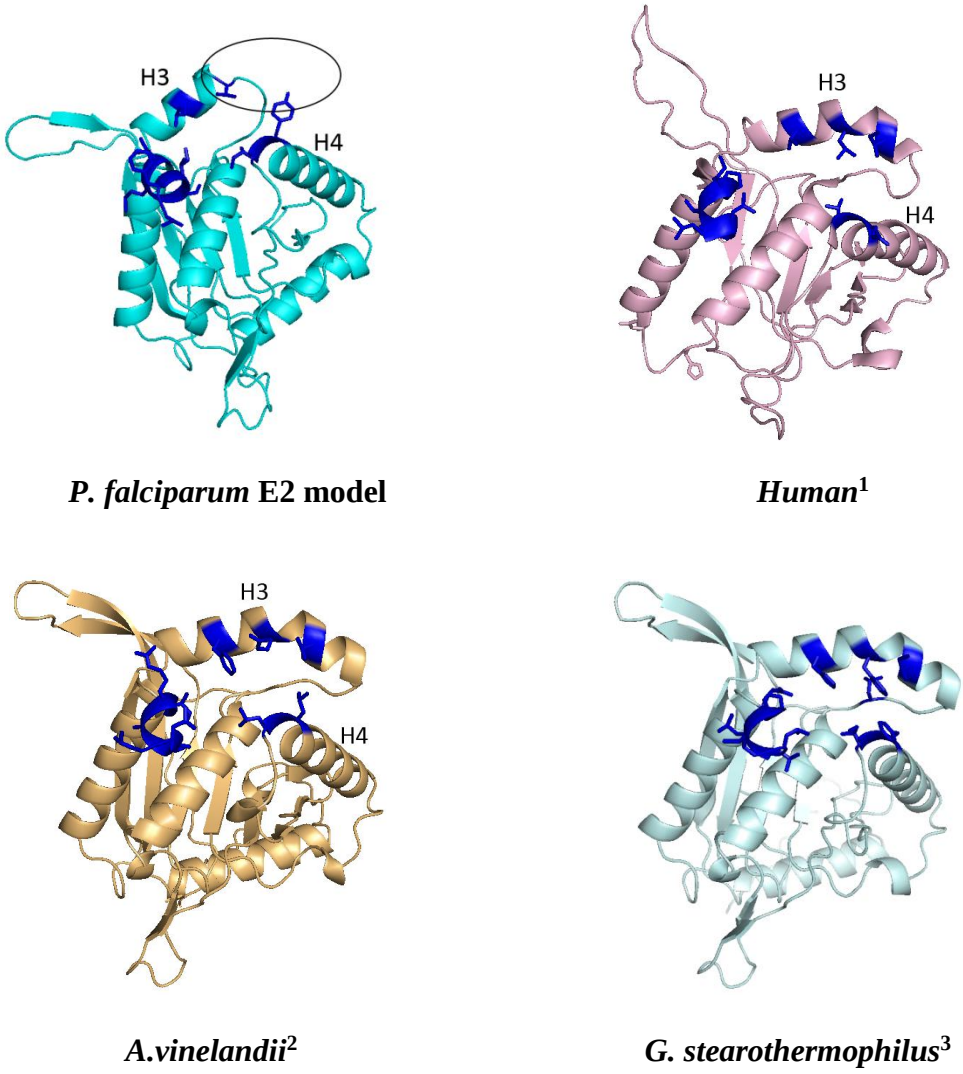


Figure 4-7. Potential model of interaction between two *Pf*E2 trimers
Residues involved in inter-trimer interactions based on those known to be involved in trimer formation in *A. vinelandii* and *G. strearothermophilus* (Mattevi *et al.*, 1992a; Izard *et al.*, 1999). Residues involved are in dark blue and red. (A) View of two trimers from the inner surface facing the centre of the core. (B) The residues thought to be involved in inter-trimer bonds. (C) Hydrogen bonds could be formed between Phe639 and Ile 529, and Leu 640 and Tyr530.

The required hydrophobic nature of the residues involved is conserved. In *A. vinelandii* and *G. strearothermophilus*, the last two residues in the 3_{10}-helix form hydrogen bonds with residues at the N-terminus of helix H4 (Mattevi *et al.*, 1992a; Izard *et al.*, 1999). The hydrogen bonds are not formed by the side chains, instead they are between the main chain nitrogen and carbonyl oxygen (Mattevi *et al.*, 1992a; Izard *et al.*, 1999). Thus, the corresponding residues in *P. falciparum*, Phe639 and Leu640 at the C-terminus and Ile529 and Tyr 530 in helix H4, could also form these bonds (Fig 4-7C).

When comparing the sites involved in the inter-trimer interactions to the bacterial and human counterparts, the *P. falciparum* site is missing a residue, as helix H3 is shorter (Fig. 4-8). The residues involved in the latter part of H3 in *A. vinelandii*, *G. strearothermophilus* and human CD structures are absent in *P. falciparum* as seen by both the ClustalW2 alignments (Fig. 3-4) and I-TASSER alignment of *Pf*CD with template structures. Interestingly, the ClustalW2 alignment (Fig. 3-4) also shows that *A. thaliana* chloroplast and *Synechocystis* sp. CD amino acid sequences also lack these residues.



***P. falciparum* E2 model** ***Human*[1]**

***A.vinelandii*[2]** ***G. stearothermophilus*[3]**

Figure 4-8. Comparison of inter-trimer interaction sites in different organisms
The *P. falciparum* helix H3 is shorter than that from other organisms, the missing part of the helix is circled. Residues involved in inter-trimer interactions in dark blue. [1]PDB entry 3B8K, Yu *et al.* (2008b); [2]PDB entry IEAE, Mattevi *et al.* (1993a); [3]PDB entry 1B5S, Izard *et al.* (1999).

Attempts were also made to model the protein in sections. Modelling the two LDs with their joining linker, and LD2 connected to the SBD were not successful. However, a model was obtained for the SBD linked to the CD. In the model the linker of one monomer reaches over the neighbouring monomer, just as in the CD homology model. The SBD then reaches back on top of its own monomer (Fig. 4-9). The residues in the linker, which may be involved in the trimer formation, are positioned the same way as in the CD model (Fig 4-10). This arranges the SBD in an orientation where the residues that potentially interact with E1 and E3 face the surface of the trimer and would not be accessible to the other proteins. Movement from the linker would be required to orient the SBD correctly, however, there are only 8 – 10 residues between the SBD and the first residue in the linker,

which possibly interacts with the CD in trimer formation. The residues involved in trimer formation are based on crystal structures, thus it could be that in solution these interactions do not occur, or are transient.

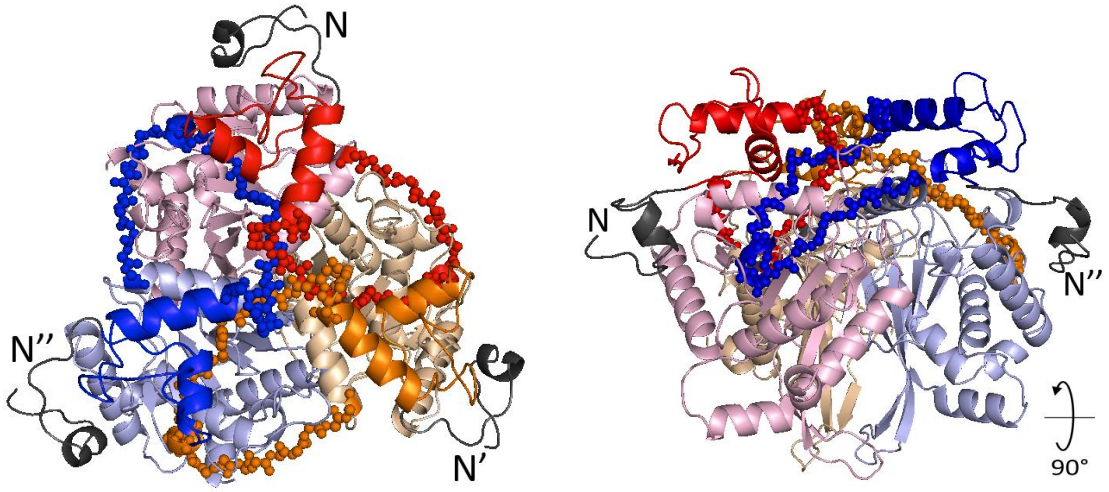


Figure 4-9. Homology models of the SBD linked to the CD
Each monomer is shown in a light colour (pink, light orange, light blue). The corresponding linker and SBD are shown in red, orange and blue. The linker is shown as spheres. The His-tag of the His-r*Pf*E2bc construct is in grey.

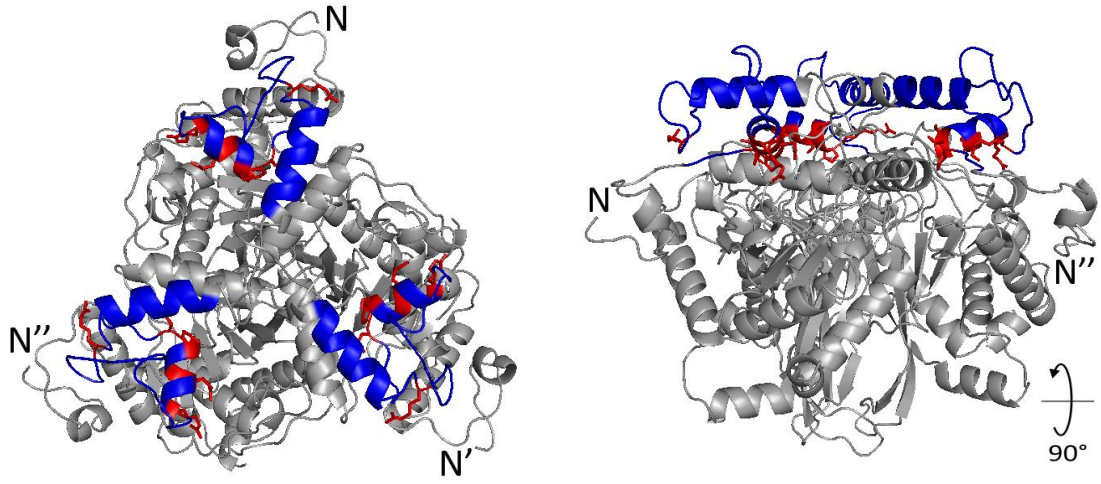


Figure 4-10. Orientation of the SBD in the *Pf*E2 homolgy model
The SBD is shown in blue. The residues and their side-chains that are potentially involved in E1/E3 binding are in red.

4.1.3.2 Stoichiometryof His-r*Pf*E2m

4.1.3.2.1 Sedimentation coefficient of His-r*Pf*E2m species purified by gel filtration

To determine the stoichiometry of His-r*Pf*E2m, AUC was carried out on each of the four species of His-r*Pf*E2m purified by gel filtration (Fig. 3-15C). The reference buffer used was PEBS100 (Table 2.9) and the samples were initially spun at 49,000 rpm as the molecular masses of the species in the samples were unknown. This speed was appropriate for His-r*Pf*E2m from peaks 3 and 4. However, peaks 1 and 2 comprised larger molecular mass species, thus in subsequent AUC analyses the samples were centrifuged at 3,000 rpm. For SEDFIT (Schuck, 2000; Schuck *et al.*, 2002) analysis the following parameters were used: His-r*Pf*E2m $\bar{\nu}$ = 0.7338 ml/g at 4ºC and 0.7406 ml/g at 20ºC; PEBS100 ρ = 1.01188 g/ml and η = 1.6141 cPoise (Table 2-8 and 2-9).

His-r*Pf*E2m which eluted at ~100 ml (peak 1) on the HiPrep 26/60 Sephacryl S-400 HR gel filtration column (Fig. 3-15C), was expected to have a large molecular mass, possibly corresponding to a large E2 core. Protein from two separate purifications was shown to contain very large species; one sample contained a single species with $s_{20,w}$ 100 S, the other comprised three species corresponding to 40 S, 100 S and 130 S (Fig 4-11). Unfortunately, these values are much larger than the sedimentation coefficient expected for an E2 core, as the sedimentation coefficient of a 24-mer core is 23 S (Willms *et al.*, 1967) and of a 60-mer is ~30 S (Vijayakrishnan *et al.*, 2010). The finding that peak 1 contained mostly aggregates was supported by the observation that, when determining protein concentration spectrophotometrically by absorbance at 280 nm, much larger values (0.7 – 3.5 mg/ml) would be obtained than when using the Bradford method (0.2-0.8 mg/ml). Also the amount of protein seen on Coomassie blue or silver stained SDS-PAGE gels did not correspond to high protein concentrations. Large aggregates in a sample would scatter light and cause apparent higher absorbance values at 280 nm.

His-r*Pf*E2m in peak 2, which eluted at ~130 ml from the HiPrep 26/60 Sephacryl S-400 HR gel filtration column, was always at a very low concentration. SV analysis on two separate occasions showed the samples were also mainly aggregated protein (Fig. 4-11).

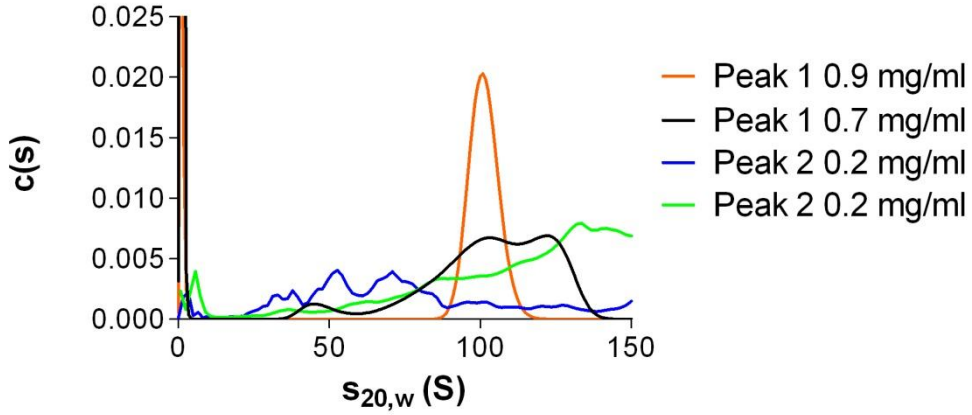


Figure 4-11. SV analysis of His-r*Pf*E2m peak 1 and 2
c(*s*) analysis of the SV data was carried out using SEDFIT (Schuck, 2000; Schuck *et al.*, 2002). Each sample was from a separate purification carried out 24 h prior to the start of the SV experiment. Only protein in peak 1 formed distinct species, however, $s_{20,w}$ values were extremely large, indicating that these samples contained only aggregated His-r*Pf*E2m. Interpretation of the data for peak 2 was difficult, as the SV data could not be well described by distinct species even beyond $s_{20,w}$ 150 S. The data was analysed up to $s_{20,w}$ 200 S.

His-r*Pf*E2m in peak 3 (eluted at ~180 ml) or 4 (eluted at ~220 ml) did not form large complexes either. Peak 3 contained a species with $s_{20,w}$ 6.0 S while peak 4, consisted of approximately equal amounts of two species with $s_{20,w}$ 3.9 S and 6.1 S (Fig. 4-12). Because most of the active His-r*Pf*E2m was found in peak 3 following purification, and the SV analysis showed the protein in this peak was also the purest, His-r*Pf*E2m in peak 3 was taken as the representative sample of the protein for further analysis.

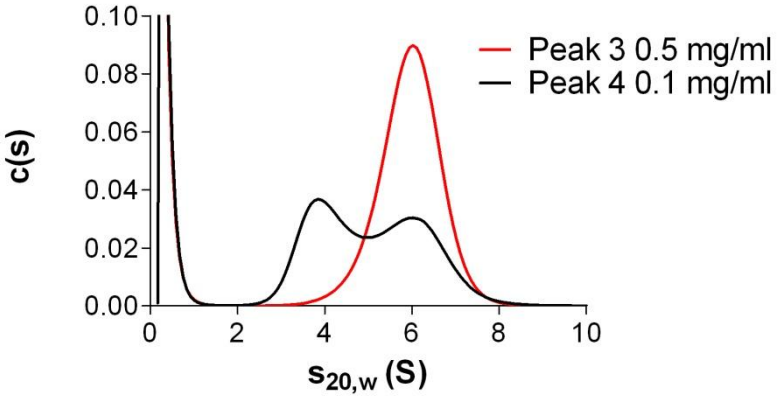


Figure 4-12. SV analysis of His-r*Pf*E2m peak 3 and 4
c(*s*) analysis of the SV data was carried out using SEDFIT (Schuck, 2000; Schuck *et al.*, 2002). His-r*Pf*E2m peak 3 was homogenous whereas peak 4 contained roughly equal amounts of two species. Due to the very low concentration of peak 4 and thus, low absorbance at 280 nm, for this experiment only the absorbance optics were set at a wavelength of 225 nm.

4.1.3.2.2 The effect of buffer NaCl concentration on His-r*Pf*E2m

As initial SV analyses showed only aggregates and no large complex, experimental conditions which could affect the assembly of the E2 core were considered. NaCl is added to purification buffers to inhibit non-specific ionic interactions between proteins themselves, and also interactions between proteins and purification resins. As the interactions between the E2 trimers to form the large core are relatively weak, the effect of NaCl concentration on the stoichiometry of His-r*Pf*E2m was tested. Following gel filtration, His-r*Pf*E2m from peak 3 was dialysed against buffers with NaCl conentrations ranging from 0 to 1 M. These samples were then analysed by SV to observe the effect on the species present and the heterogeneity of the samples.

SV was carried out at 49,000 rpm. The density and viscosity values for each of the buffers are listed in Table 2-9. NaCl concentration did not affect the formation of larger His-r*Pf*E2m complexes. SV analysis confirmed the presence of a main species of $s_{20,w}$ = 5.9 – 6.6 S and a minor species of $s_{20,w}$ = 3.2 – 3.6 S (Fig 4-13). NaCl concentration did, however, have an effect on sample heterogeneity, which was taken into account for subsequent purifications. In buffers with 100, 250 and 500 mM NaCl, 63, 66 and 68% of the protein was found in the ~6.3 S peak, respectively, compared with only 38 – 49% with 0 – 50 mM NaCl (Table 4-3). These samples also showed least aggregation (26-27% of sample) compared with the amount of aggregation seen with lower NaCl concentrations (40 - 52%). In 1 M NaCl, 60% of His-r*Pf*E2m was in the ~6.3 S peak, but also more aggregation (36% of sample) could be seen. The amount of the minor ~ 3.4 S species stayed fairly constant (10 – 15%) in 0 – 100 mM NaCl, with slightly less (4-8%) found at higher NaCl concentrations. The sedimentation coefficient for His-r*Pf*E2m in 0 – 250 mM NaCl was stable (6.4 – 6.6 S), while it decreased to 6.2 S in 500 mM and 5.9 S in 1 M NaCl. To keep the salt concentration at a level suitable for downstream applications (e.g. SAXS) while minimising heterogeneity of the sample and keeping the sedimentation coefficient in the stable range, buffer with 250 mM NaCl was used hereafter.

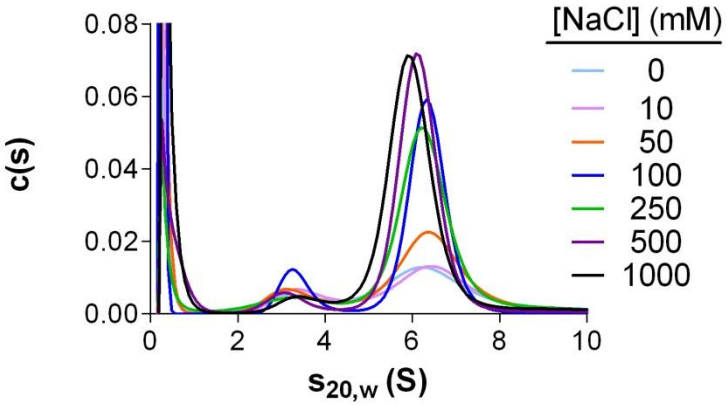


Figure 4-13. SV analysis of His-r*Pf*E2m at various NaCl concentrations
c(*s*) analysis of the SV data was carried out using SEDFIT (Schuck, 2000; Schuck *et al.*, 2002).All samples showed the presence of a main peak (~6.3 S) and a minor peak (~3.4 S).

				% of total protein sample		
[NaCl] (mM)	[His-r*Pf*E2m] (μM)	$s_{20,w}$ (S)		~3.4 S	~6.3 S	Aggregate
0	1.9	3.4	6.4	10	38	52
10	2.0	3.5	6.6	15	42	43
50	1.4	3.4	6.5	11	49	40
100	1.9	3.3	6.4	10	63	27
250	1.4	3.2	6.4	8	66	26
500	1.4	3.2	6.2	6	68	26
1000	1.4	3.6	5.9	4	60	36

Table 4-3. Percentage of each species present in His-r*Pf*E2m at various NaCl concentrations

4.1.3.2.3 SV analysis of various concentrations of His-r*Pf*E2m

SV of His-r*Pf*E2m over a concentration range of 4 – 17 μM was carried out with PEBS250 (ρ = 1.01804 g/ml and η = 1.6354 cPoise; for buffer composition see table 2.1) as reference buffer. A main species (33 – 40% of sample) with $s_{20,w}$ 6.6 – 7.0 S and a minor 10.2 – 10.6 S species (4-8% of sample) could be observed (Fig. 4-14A, Table 4-4). At the lowest concentration (4 μM) a 2.5 S species (24% of sample) could be seen, which was most likely a degradation product. Regardless of optimising the sample buffer, aggregation (species with sedimentation coefficients over 40 S) could still be seen in increasing amounts as protein concentration was increased (Table 4-4). At the highest concentration (17 μM) a total of 68% of the sample was aggregated, while at 4 μM the percentage was only 32%. The $s_{20,w}$ values were plotted against His-r*Pf*E2m concentration to obtain an infinite dilution sedimentation coefficient, which was $s^{0}_{20,w}$ 6.6 ± 0.1 S (Fig 4-14B).

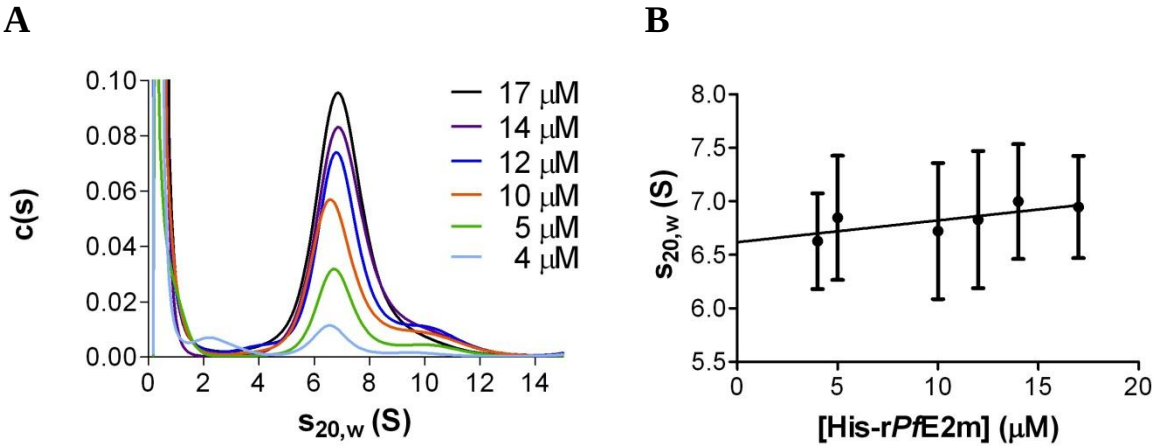


Figure 4-14. SV analysis of various concentrations of His-r*Pf*E2m
c(*s*) analysis of the SV data was carried out using SEDFIT (Schuck, 2000; Schuck *et al.*, 2002). (A) SV analysis of six different concentrations (4 – 17 μM) of His-r*Pf*E2m show the presence of a main species with an average $s_{20,w}$ 6.8 S and a minor ~10.4 S species. (B) His-r*Pf*E2m $s_{20,w}$ for the main species was plotted against concentration to obtain the sedimentation coefficient at infinite dilution, which was $s^{0}_{20,w}$ 6.6 ± 0.1 S. The experiment was carried out only once, thus, the error bars are the standard deviation shown as the square root of the central moment of the curve.

				% of total protein sample			
His-r*Pf*E2m (μM)	$s_{20,w}$ (S)			2.5 S	~ 6.8 S	~10.4 S	Aggregate
17	-	6.9			33		67
14	-	7.0	10.3		33	4	63
12	-	6.8	10.4		39	6	55
10	-	6.7	10.5		40	6	54
5	-	6.8	10.6		39	7	54
4	2.5	6.6	10.2	24	37	8	31

Table 4-4. Proportion of each of species present at a range of His-r*Pf*E2m concentrations from c(*s*) analysis (Fig. 4-14)

4.1.3.2.4 SE analysis of His-r*Pf*E2m

For sedimentation equlibrium analysis, various concentrations (4-17 μM) of His-r*Pf*E2m were centrifuged at 12,000 rpm. The speed was chosen based on the predicted trimer size (210 kDa) of His-r*Pf*E2m, as the major species in the SV analyses was anticipated to be a trimer. Ten scans were taken every 3 h until the sample had reached equilibrium. The non-interacting species analysis model in SEDPHAT (Vistica *et al.*, 2004) was used to analyse the SE data. Data at each concentration were analysed separately. Analysis with a single species model resulted in very poor fits at all concentrations (Fig. 4-15A) indicating the presence of more than one species in the samples. The fits were improved when two species were modelled (Fig. 4-15B).

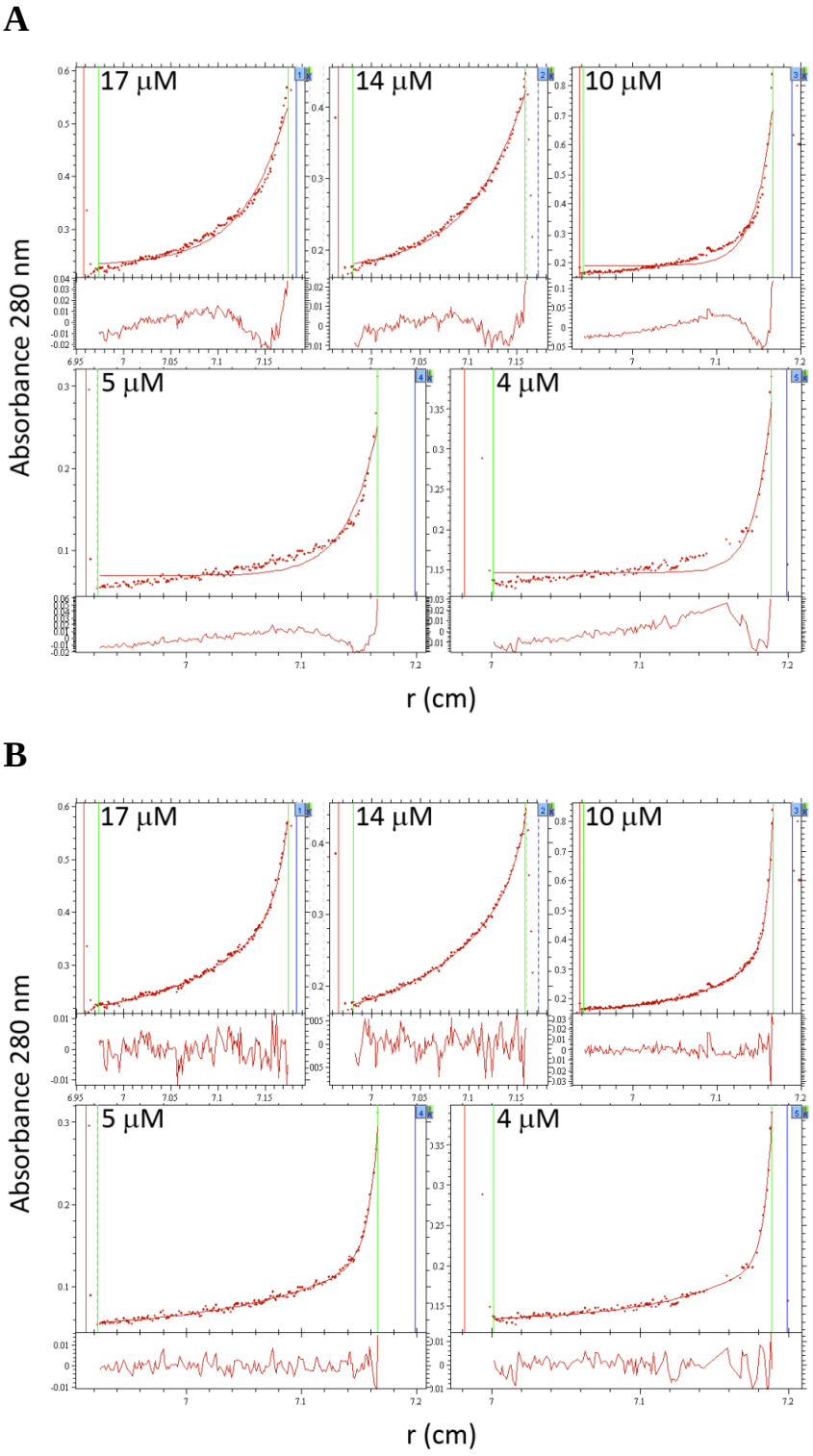


Figure 4-15. Analysis of His-r*Pf*E2m SE data
Absorbance data (dots) were fit by SEDPHAT (Vistica *et al.*, 2004) using the non-interacting discrete species model. (A) Single species analysis resulted in poor fits (smooth line) with high Chi squared (1.3-10.9) and rmsd (0.005-0.165) values. (B) A two species analysis resulted in better fits (smooth line) with lower Chi squared (0.3-0.8) and rmsd (0.003-0.004) values. The red vertical line indicates the cell meniscus, the blue vertical line indicates the cell base and the green vertical lines denote the range of data that was analysed. Residual plots, which show the goodness of the fit, are shown beneath each graph for each concentration.

The highest concentration samples, 14 μM and 17 μM, contained a 193-195 kDa species, which accounted for 84-89% of the sample (Table 4-5). This species would correspond to a trimer of His-r*Pf*E2m. The molecular mass of His-r*Pf*E2m calculated from its amino acid sequence is 210 kDa, which is slightly higher than the experimental value, suggesting some degradation of the protein.

His-r*Pf*E2m (μM)	MM	% of sample
17	195 kDa	84
	1.53 MDa	16
14	193 kDa	89
	1.33 MDa	11
10	253 kDa	28
	2.07 MDa	72
5	161 kDa	18
	2.01 MDa	82
4	287 kDa	42
	3.70 MDa	58

Table 4-5. Molecular mass (MM) of His-r*Pf*E2m at various concentrations

Interestingly, a larger 1.33 – 1.53 MDa species was also present making up 11-16% of the sample. A 24-mer of undegraded His-r*Pf*E2m would have a molecular mass of 1.67 MDa. However, taking into account the degraded His-r*Pf*E2m trimer mass of 193-195 kDa, a 24-mer would be 1.54-1.56 MDa. These values fit well with the larger molecular mass species seen in the sample, suggesting that the larger species in the samples could be a 24-mer E2 core complex. In SV experiments, however, no species with a sedimentaion coefficient that would correspond to the 24-mer size was seen. The *E. coli* E2 forms a 1.59 MDa 24-mer with a sedimentation coefficient of 23 S (Willms *et al.*, 1967), thus a species with a similar sedimentation coefficient should have been seen.

With decreasing sample concentration the molecular mass of the species increased and the main species in the samples had molecular masses of ~ 2 MDa and 3.7 MDa (Table 4-5). A 2 MDa complex would correspond to about 30 His-r*Pf*E2 monomers and 3.7 MDa would

be roughly a 60-mer. Again, species this size were not seen in SV analysis. The increase in molecular mass with decreasing concentration was evident when determining the apparent whole-cell weight average molecular mass ($M_{w,app}$). This was used to obtain the whole-cell weight average molecular mass at infinite dilution, M_w^0, by plotting sample concentration vs $M_{w,app}$ (Fig.4-16). M_w^0 for His-r*Pf*E2m was 1.54 ± 0.32 MDa.

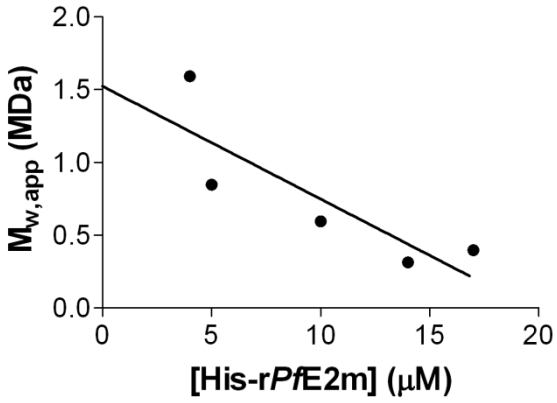


Figure 4-16. Whole cell weight average molecular mass, M_w^0, of His-r*Pf*E2m
$M_{w,app}$ was plotted as a function of sample concentration and extrapolated to zero to obtain M_w^0, which was 1.54 ± 0.32 MDa.

The increase in molecular mass with decreasing concentration is indicative of thermodynamic non-ideality of the protein in the sample. In this case, the non-ideality is most likely due to the elongated shape of His-r*Pf*E2m.

Because no 1.5 – 3.7 MDa species were observed in the SV experiments, an attempt was made to analyse the data with SEDPHAT (Vistica *et al.*, 2004) using the species analysis with mass conservation constraints (Fig. 4-17). As the presence of a trimer is clear from both SE and SV analysis, this model was used to fix the trimer size (210 kDa) and amount (based on SV analysis) to try and fit the larger species better. The analysis, however, resulted in poor fits as is evident from the residual plots (Fig. 4-17) and Chi squared (0.7-7.8) and rmsd (0.004-0.014) values. As seen in the SV analysis, the His-r*Pf*E2m sample contained a lot of aggregates (sedimentation coefficients above 40 S), which are likely to complicate the interpretation of the SE data. Thus, further analyses on this set of data were not carried out.

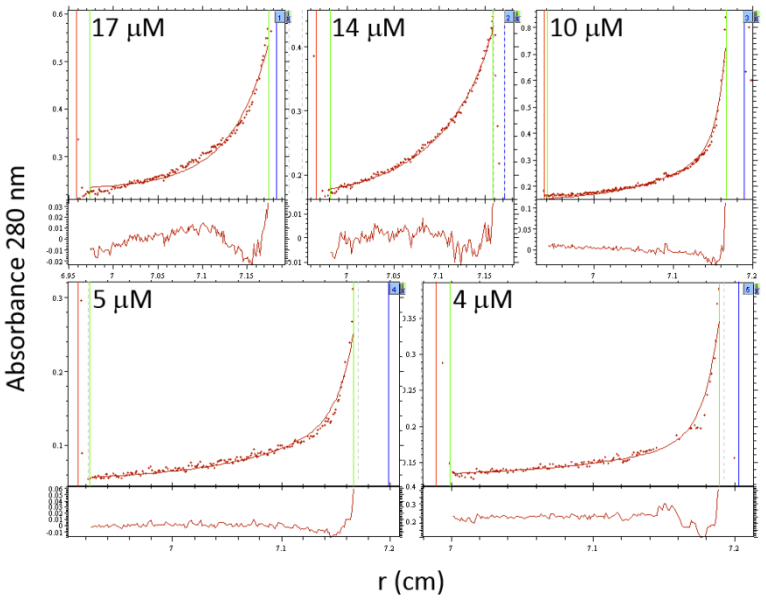


Figure 4-17. Further analysis of His-r*Pf*E2m larger molecular mass species
Absorbance data (dots) were fit by SEDPHAT (Vistica *et al.*, 2004) using a species analysis with mass conservation constrains model to obtain the fits (smooth line). The red vertical line indicates the cell meniscus, the blue vertical line indicates the cell base and the green vertical lines denote the range of data that was analysed. Residual plots are shown beneath each graph.

4.1.3.3 Small angle X-ray scattering (SAXS) of His-r*Pf*E2m

SAXS was attempted on His-r*Pf*E2m peak 1 and 3. Unfortunately, the scattering curves showed a steep increase at low angles (Fig 4-18, red circles), which is indicative of aggregation of the sample. Thus, no meaningful solution structure could be obtained for His-r*Pf*E2m.

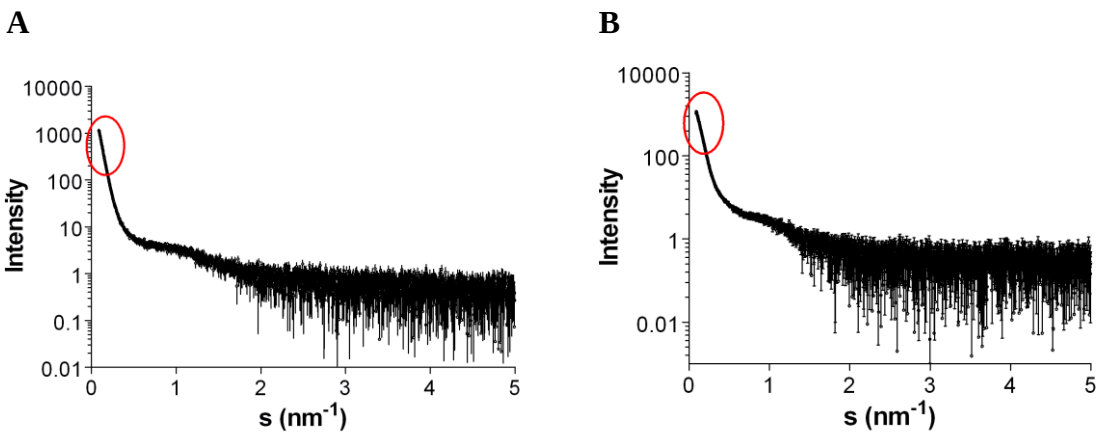


Figure 4-18. Scattering curves for His-r*Pf*E2m
(A) Peak 1 and (B) peak 2. Scattering intensity is plotted against the scattering angle, s.

4.1.3.4 Stoichiometry of His-r*Pf*E2bc

4.1.3.4.1 SV analysis of His-r*Pf*E2bc

Both of the His-r*Pf*E2bc forms purified by gel filtration (Fig. 3-16D) were analysed by AUC. The reference buffer used was PEBS150 (Table 2-1). The partial specific volumes of His-r*Pf*E2bc used in SEDFIT (Schuck, 2000; Schuck *et al.*, 2002) analysis were 0.7356 ml/g at 4ºC and 0.7424 ml/g at 20ºC. PEBS150 parameters ρ = 1.01804 g/ml and η = 1.6354 cPoise were used (Table 2-8 and 2-9).

SV at various concentrations of His-r*Pf*E2bc from gel filtration peak 1 (1 – 4 μM) showed the samples to be homogenous with a single species of $s_{20,w}$ = 4.7 – 5.2 S (Fig. 4-19A). The protein was stable at these low concentrations, as no aggregates were seen in the sample. As already mentioned previously (section 3.1.3.4.2), the protein precipitated completely at concentrations above 7-8 μM, hence no analyses could be carried out at higher concentrations. From the y intercept of a plot of His-r*Pf*E2bc concentration vs $s_{20,w}$, the infinite dilution sedimentation coefficient, $s^0_{20,w}$, was determined to be 4.5 ± 0.3 S (Fig 4-19B).

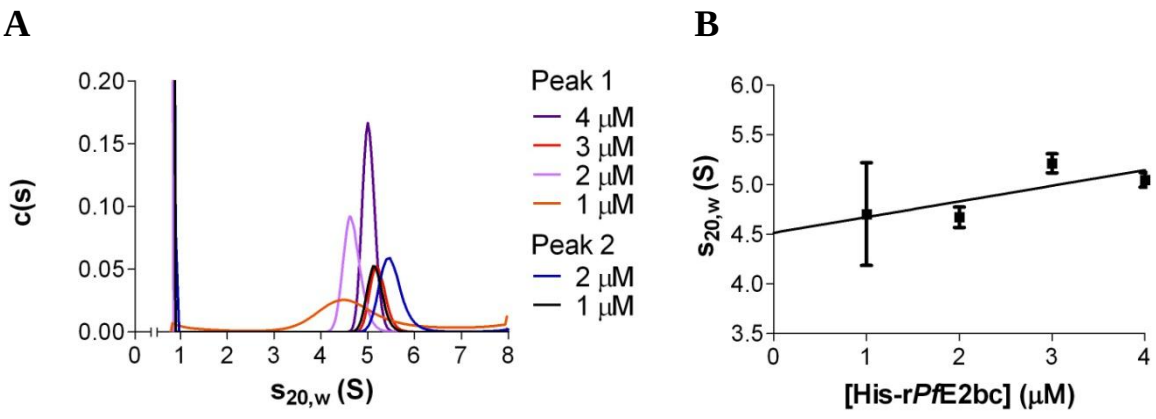


Figure 4-19. SV analysis for both forms of His-r*Pf*E2bc purified by gel filtration
(A) c(*s*) analysis with SEDFIT (Schuck, 2000; Schuck *et al.*, 2002) for His-r*Pf*E2bc from gel filtration peak 1 and peak 2. Only one species with an average $s_{20,w}$ 5.1 S was observed. (B) The $s_{20,w}$ of His-r*Pf*E2bc from peak 1 was plotted against concentration to obtain the infinite dilution sedimentation coefficient, which was $s^0_{20,w}$ = 4.5 ± 0.3 S. The experiment was carried out only once, thus, the error bars are the standard deviation shown as the square root of the central moment of the curve.

His-r*Pf*E2bc from gel filtration peak 2 had sedimentation coefficients ($s_{20,w}$ = 5.2 S and 5.5 S) similar to those for peak 1 (Fig. 4-18A). Due to the very low concentration of this peak, only two samples of 2 μM and 1 μM were analysed, and therefore no $s^0_{20,w}$ value could be determined.

4.1.3.4.2 SE analysis of His-r*Pf*E2bc

For SE analysis of His-r*Pf*E2bc from peak 1, two concentrations (4 μM and 2 μM) were centrifuged at 35,000 rpm (based on trimer His-r*Pf*E2bc size, 104 kDa) and 10 scans were taken every 3 h until equilibrium was reached. The non-interacting species analysis model in SEDPHAT (Vistica *et al.*, 2004) was used to analyse the SE data. Good fits were obtained using a single species model resulting in molecular masses of 103 kDa and 108 kDa for the 4 μM and 2 μM samples respectively (Fig. 4-20). This compared well with the molecular mass of 104 kDa calculated from the amino acid sequence for a His-r*Pf*E2bc trimer. Two samples of His-r*Pf*E2bc from peak 2 (2 μM and 1 μM) were also analysed, however, unfortunately the samples had degraded during the experiment.

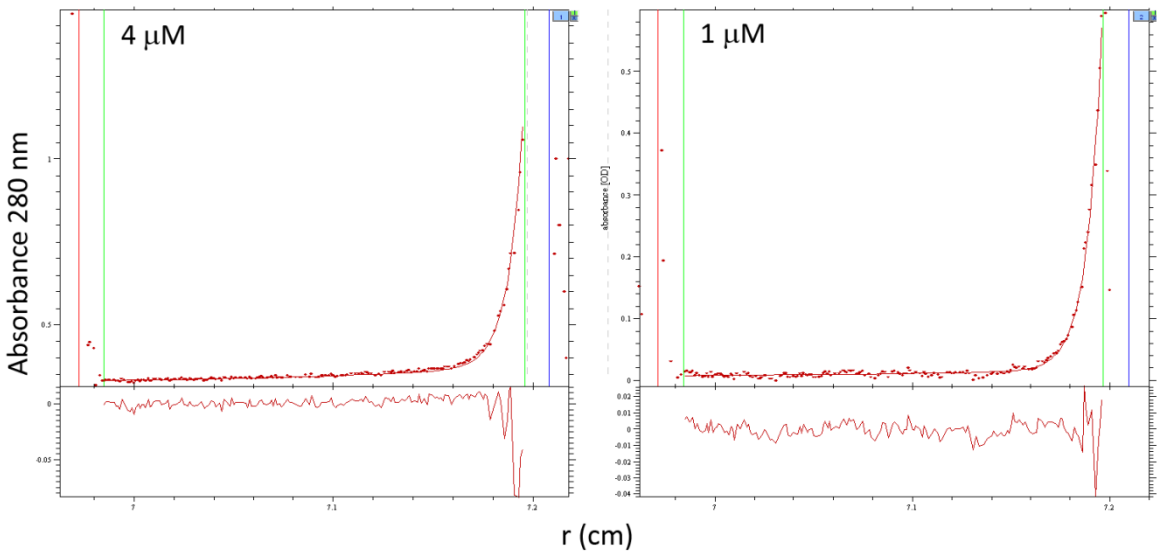


Figure 4-20. Analysis of His-r*Pf*E2bc SE data
Absorbance data (dots) were fit by with SEDPHAT (Vistica *et al.*, 2004) using the non-interacting discrete single species model (smooth line). The red vertical line indicates the cell meniscus, the blue vertical line indicates the cell base and the green vertical lines denote the range of data that was analysed. Residual plots are shown beneath each graph for each concentration.

4.1.3.5 Solution structure of His-r*Pf*E2bc

4.1.3.5.1 Small-angle scattering of His-r*Pf*E2bc

SAXS measurements were carried out at the EMBL/DESY beamline X33. Homogeneity of protein samples was determined by an SV experiment before transport to the beamline. His-r*Pf*E2bc in peak 1 consisted of a 4.8 S main species (88%, Fig. 4-21). A minor, 1.2 S species (12%) was also detected, which could have been a degradation product. Peak 2 comprised two species, the 4.7 S major component (76%) and a 2.8 S species (24%, Fig. 4-21).

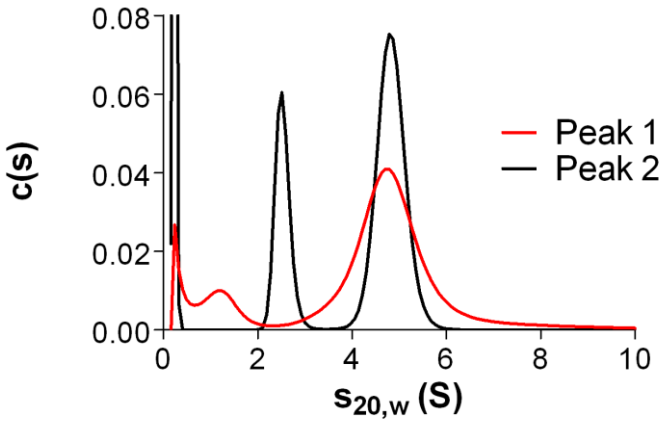


Figure 4-21. SV analysis of His-r*Pf*E2bc samples prior to SAXS
The concentration of both samples was 3 μM. His-r*Pf*E2bc in peak 1 was mostly homogenous, while peak 2 contained two species.

Due to the instability of His-r*Pf*E2bc at high concentrations, the samples were transported to the beamline at very low concentrations (0.4 – 0.5 μM). The final concentrations of peak 1 and 2 His-r*Pf*E2bc used for SAXS measurements were 5 μM and 3 μM, respectively. As the concentration of the samples was still low, the data were very noisy and detail in the scattering data at high angles is very limited (Fig. 4-22), however, the data from low angles were of sufficient quality to provide some useful structural information.

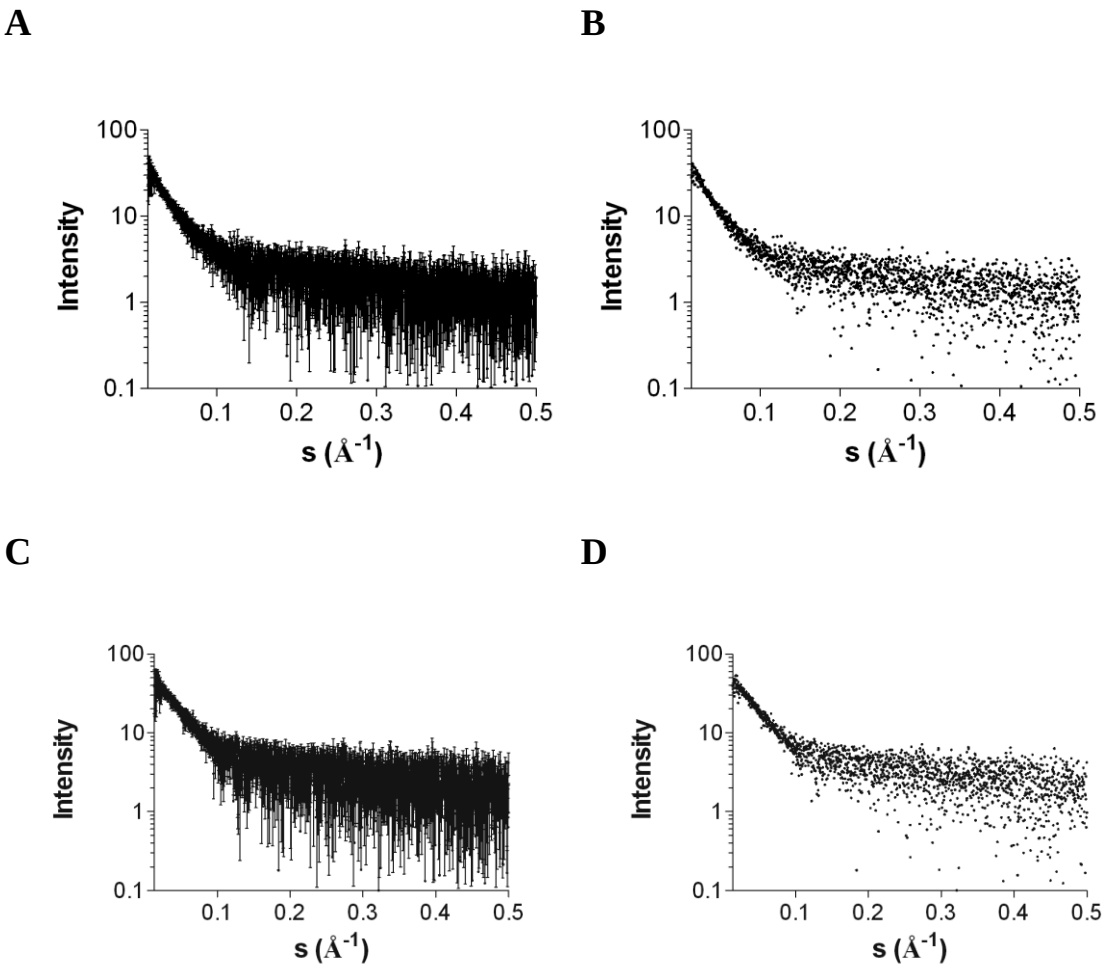


Figure 4-22. SAXS curves for His-r*Pf*E2bc
(A) Curve for His-r*Pf*E2bc in peak 1 with error bars and (B) without error bars. (C) Curve for His-r*Pf*E2bc in peak 1 with error bars and (D) without error bars.

The linear Guinier region confirmed that there was no aggregation in the sample (Fig. 4-23A & B) and the radius of gyration, R_g, was determined to be 43.1 ± 0.4 Å and 33.7 ± 0.2 Å for peak 1 and 2, respectively. GNOM (Svergun, 1992) was used to determine the distance distribution function, *p(r)*, from which the maximum dimension of the particle, D_{max}, was determined to be 147 Å for peak 1 and 113 Å for peak 2 (Fig. 4-23C & D). The values of R_g estimated from GNOM were similar to those determined from the Guinier region; 45.2 ± 0.4 Å for peak 1 and 34.8 ± 0.5 Å for peak 2. The shape of the *p(r)* curve is suggestive of a particle with separated subunits (Fig. 1-21) as would be the case with His-r*Pf*E2bc.

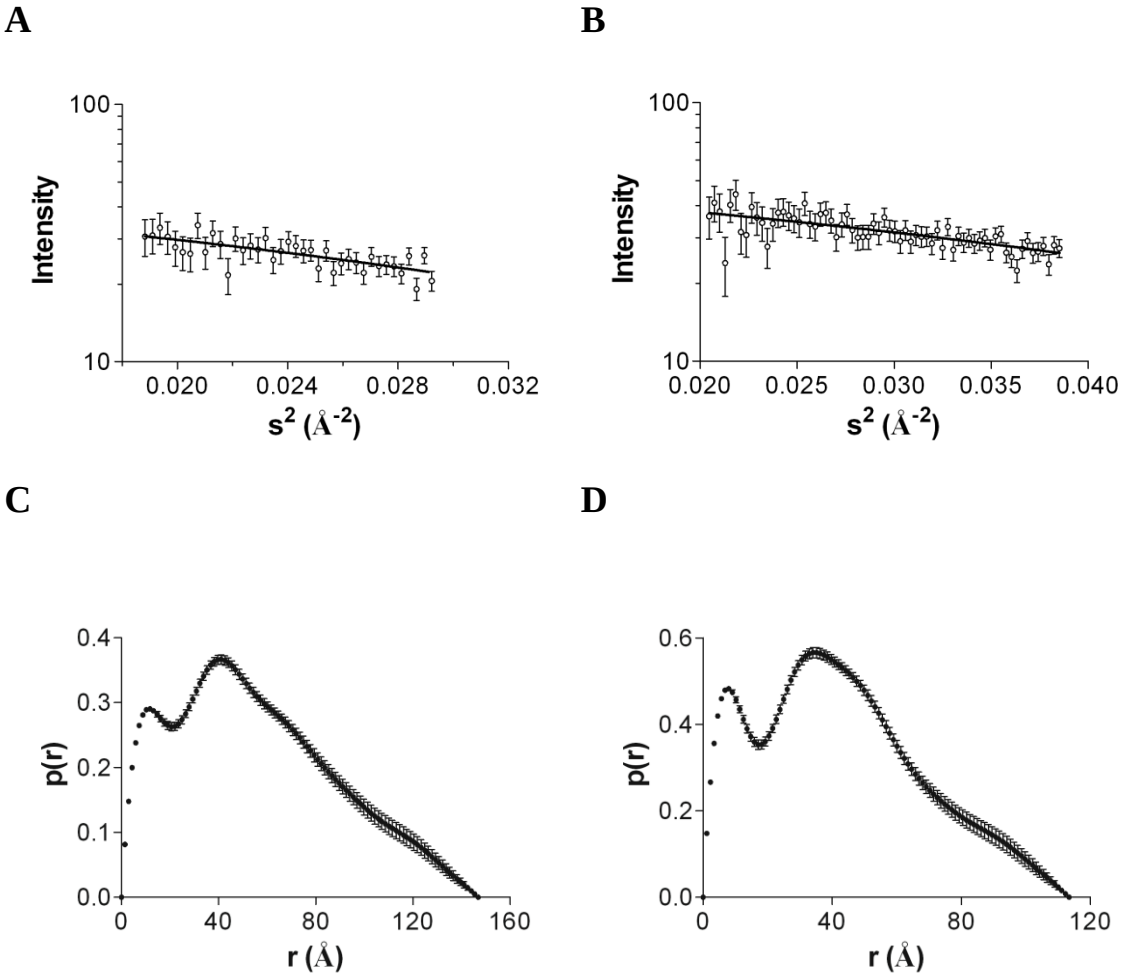


Figure 4-23. SAXS data analysis for His-r*Pf*E2bc
Guinier analysis of SAXS data aquired for His-r*Pf*E2bc peak 1 (A) and peak 2 (B). The R_g was 43.1 ± 0.4 Å for peak 1 and 33.7 ± 0.2 Å for peak 2. The *I*(0) was 38.2 ± 2.2 for peak 1 and 49.5 ± 1.8 for peak 2. Distance distribution for peak 1 (C) and peak 2 (D). The D_{max} for peak 1 was 147 Å and for peak 2, 113 Å.

The Kratky plots do not show a bell-shaped curve characteristic of folded proteins, thus His-r*Pf*E2bc could be unfolded (Fig. 4-24A & B). However, activity assays carried out on the samples before SAXS measurements showed the protein to be active, thus the protein may be merely partially unfolded. The molecular mass of His-r*Pf*E2bc estimated by extrapolating the scattering intensity to zero angle, *I(0)* was 38 kDa and 49 kDa, for peak 1 and peak 2 respectively. This would suggest the His-r*Pf*E2bc sample to contain only monomer, however, the AUC data (Fig. 4-21) confirmed that in peak 1 88% of the sample has a sedimentation coefficient consistent with a trimer.

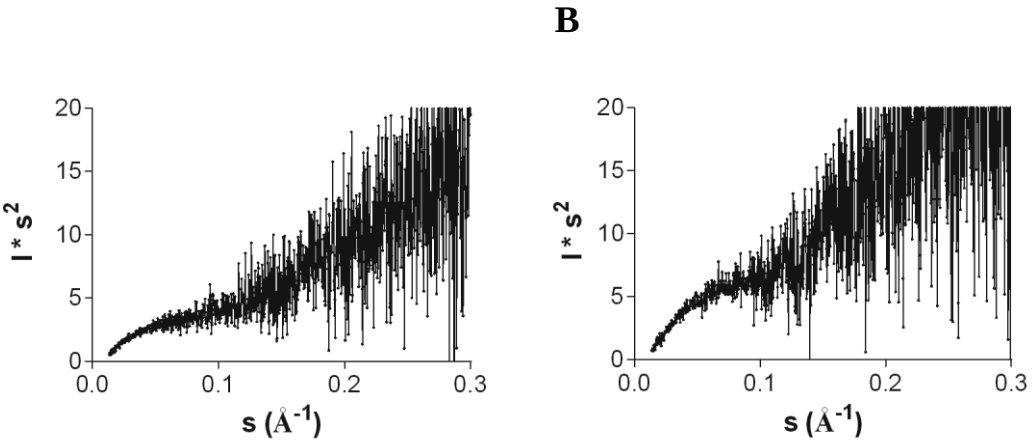


Figure 4-24. Analysis of His-r*Pf*E2bc folding
Kratky plots for peak 1 (A) and peak 2 (B) indicative of partial unfolding of His-r*Pf*E2bc.

4.1.3.5.2 *Ab initio* modelling

Ab initio modelling was attempted using DAMMIF (Franke and Svergun, 2009) on the ATSAS online server (http://www.embl-hamburg.de/biosaxs/atsas-online/dammif.php, Petoukhov *et al.*, 2007, section 2.8.3). Using P3 symmetry, 20 *ab initio* models were generated, which were then clustered by DAMCLUST (Petoukhov *et al.*, 2012). For His-r*Pf*E2bc peak 1, the models were clustered into eight groups of which seven contained a single model and one group had 13 models. The representative *ab initio* model of the large cluster is shown in Fig. 4-25A. The *ab initio* models can be further refined by averaging the models in a cluster with DAMAVER (Volkov and Svergun, 2003) and use DAMFILT (Volkov and Svergun, 2003) to obtain an averaged model (Fig. 4-25B). The model had a R_g of 45.5 Å and a D_{max} of 159 Å and fit the simulated scattering curve generated by DAMMIF well (χ^2 = 1.04, Fig. 4-26A). An optimal fit would have a Chi squared value of 1.0. The DAMMIF generated scattering curve and fit were scaled and overlaid with the actual experimental scattering data and the simulated curve looks similar to the experimental data, thus adding to the validity of the *ab initio* model (Fig. 4-26B).

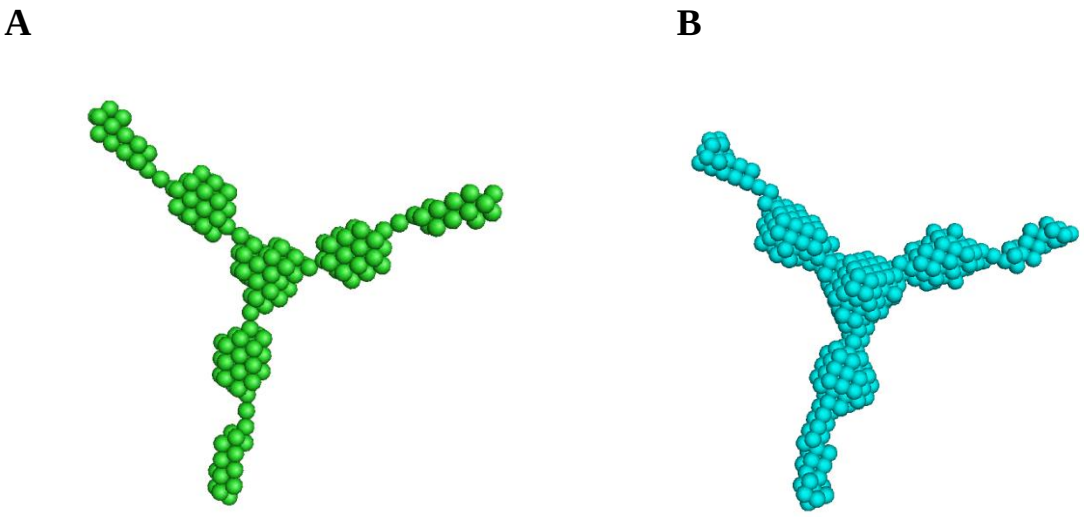


Figure 4-25. *Ab initio* model for peak 1 His-r*Pf*E2bc
(A) His-r*Pf*E2bc peak 1 representative model from the largest DAMMIF cluster. (B) The averaged model for the largest cluster.

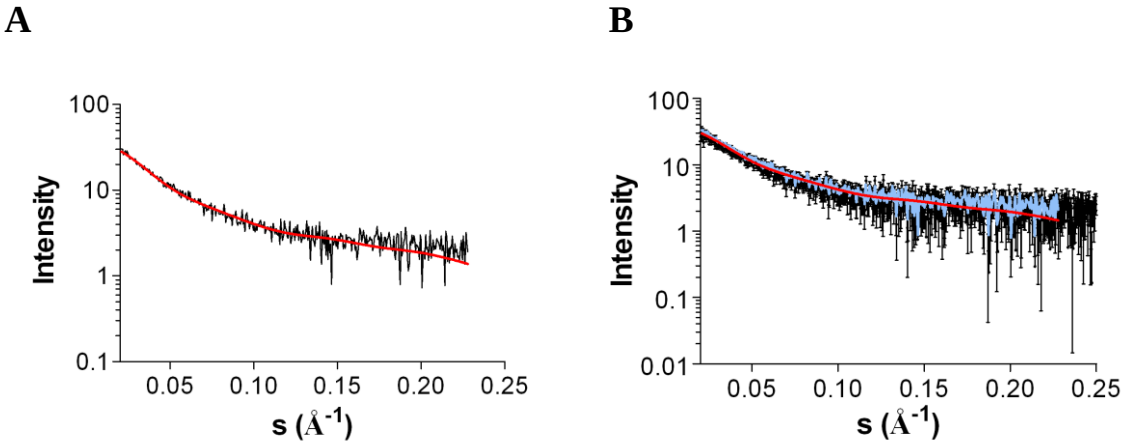


Figure 4-26. Fit of *ab initio* model for peak 1 His-r*Pf*E2bc to scattering data
(A) Fit of the representative model (red) to the simulated scattering curve (black) generated by DAMMIF. (C) Overlay of the simulated scattering curve (light blue) and the representative model fit (red) to the experimental scattering curve (black).

The models in the other clusters were all similar in that the arms of the structure were oriented on the z-axis rather than the x/y-axis. These models had on average poorer fits to the simulated scattering curve (χ^2 = 1.18) than the models in the largest cluster (average χ^2 = 1.04). However, three of the models had Chi squared values ranging between 1.03 – 1.05, a representative of these is shown in Fig. 4-27.

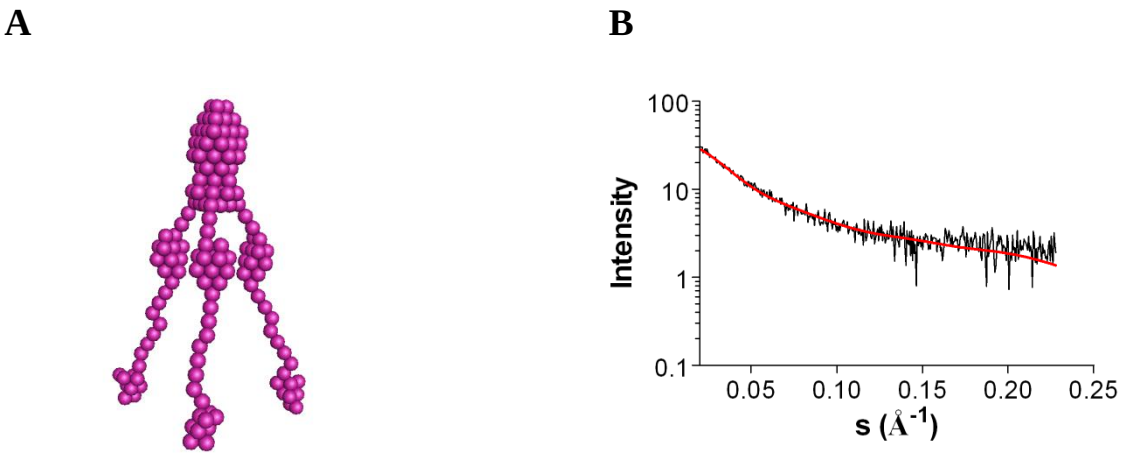


Figure 4-27. *Ab initio* model from smaller cluster
(A) One of the DAMMIF models in a smaller cluster. (B) Fit of the model (red) to the simulated scattering curve (black) (χ^2 = 1.05).

Modelling of SAXS data for His-r*Pf*E2bc peak 2 with P3 symmetry resulted in seven clusters. The largest cluster comprised only eight models, the representative model and the fit to the simulated scattering curve (χ^2 = 1.05) is shown in Fig. 4-28.

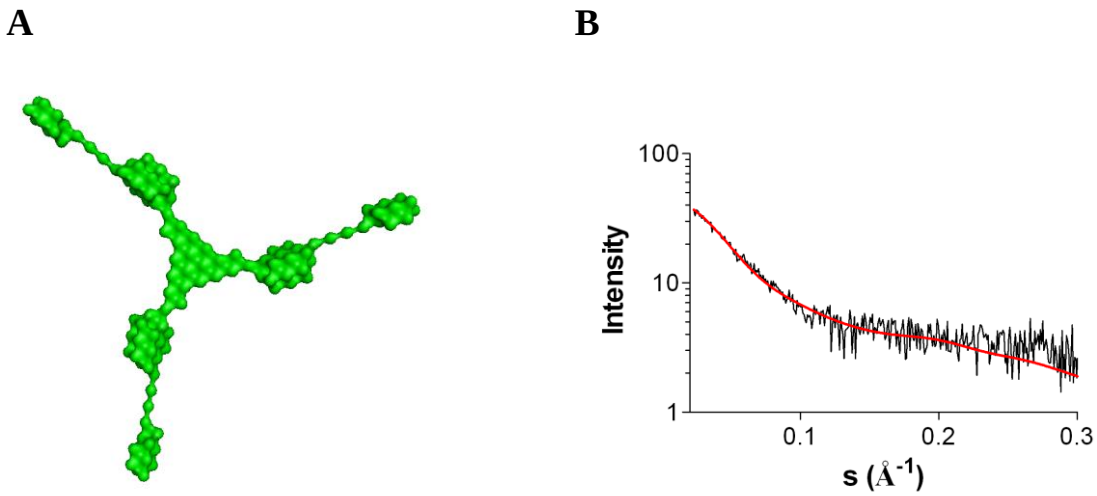


Figure 4-28. *Ab initio* model for peak 2 His-r*Pf*E2bc
(A) His-r*Pf*E2bc peak 2 representative model from the largest DAMMIF cluster. (B) Fit of the representative model (red) to the experimental scattering curve (black).

The R_g of the model was 35 Å and the D_{max} 115 Å. The models in this cluster were identical, thus no average model was generated. There was more variation in the *ab initio* models for His-r*Pf*E2bc in peak 2. The six other clusters contained models that varied not

only between the clusters but within them as well. The fits of these models to the simulated scattering curve were also worse as the Chi squared values ranged between 1.10 - 1.29. The generation of several different DAMMIF models is indicative of unfolded protein, thus it is possible that the protein in this sample had denatured during transport to the synchrotron.

To address the possibility of peak 2 His-r*Pf*E2bc consisting of both the monomer and trimer species, as suggested by the SV experiment prior to SAXS measurements, the programme OLIGOMER (Petoukhov *et al.*, 2012) was used. When providing monomer and dimer form factors, i.e. information about shape and size of the two conformations, the programme calculates the proportion of each species present in the sample. Both the I-TASSER and *ab initio* models were used to obtain these values and in both cases, only the trimer species was present. Thus, although the DAMMIF models for the protein in this sample seems suspicious and might indicate unfolded protein, the OLIGOMER result would suggest that there is an at least partially folded trimer in the sample. His-r*Pf*E2bc in peak 1 was also analysed using OLIGOMER and only trimer was present as well.

The estimated molecular masses of the *ab initio* models determined by DAMMIF were very low, around 17 kDa, which is apparent from the models being very stringy and lacking the volume of the expected catalytic and binding domains. This could be due to the very low concentration of the samples resulting in the high noise-to-data ratio, especially at high angles. It could also be due to DAMMIF not being the ideal method for modelling flexible proteins.

His-r*Pf*E2bc is flexible due to the 31 amino acid linker between the binding domain and the catalytic domain. The ensemble optimisation method (EOM) (Bernado *et al.*, 2007; Petoukhov *et al.*, 2012) is a modelling programme for flexible proteins. Given a suitable starting model, EOM generates a pool of 10,000 conformations, which cover the experimental configurational space. From this, a genetic algorithm is used to generate an ensemble of models, which fit the experimental data. High-resolution structures are used as initial possible conformations for the ensemble. In this case, the I-TASSER models were used. The catalytic domain (CD) was fixed, while the linker and sub-unit binding domain (SBD) were allowed to adopt all sterically possible conformations. The simulation was performed three times. For peak 1, ensembles containing 16, 18 and 24 models in each

EOM simulation were generated. The ensembles fit the data well at low angles, but not at high angles, most likely due to the lack of structural information caused by the low concentration of the sample (Fig. 4-29A). The Chi squared value of the overall fit for all curves was 1.03. Modelling was also tried once using monomer symmetry, however, the fit of the ensemble was worse (χ^2 = 1.09), especially at low angles (Fig. 4-29B). Thus, the presence of a trimer of His-r*Pf*E2bc was further confirmed.

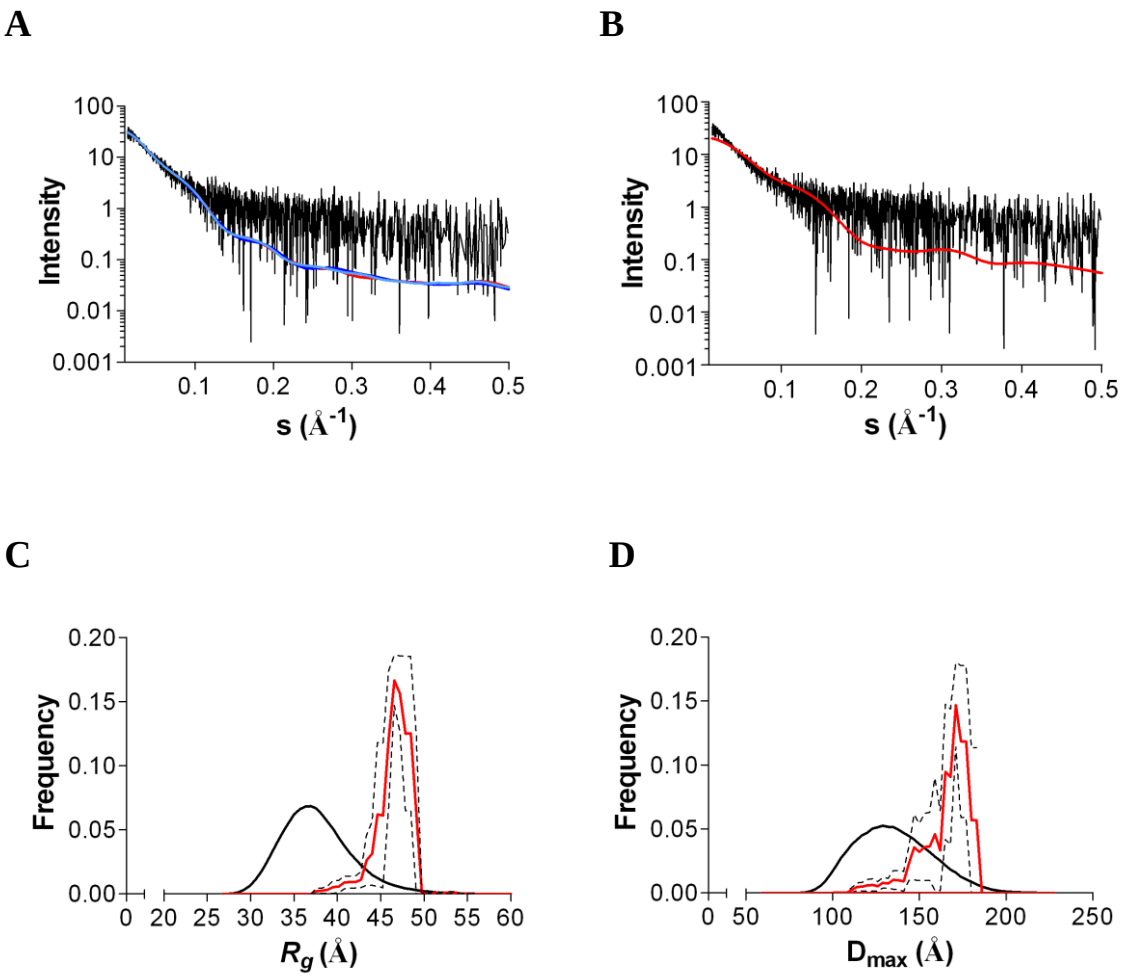


Figure 4-29. EOM of His-r*Pf*E2bc in peak 1
(A) Fit of ensemble curves from three EOM simulations using trimer symmetry (red, dark blue and light blue) to the scattering curve (black). (B) Fit of ensemble curve from one EOM simulation using monomer symmetry (red) to the scattering curve (black). (C) R_g distributions of pool (black) and the average of three EOM simulations (red), with error shown as dashed line. (D) D_{max} distributions of pool (black) and the average of three EOM simulations (red), with error shown as dashed line.

How flexible a protein is can be determined by plotting the R_g and the D_{max} values of the pool and the ensembles. If the range of ensemble values is as wide as the pool, the protein is very flexible, if the ensemble values are very constrained the protein is less flexible. The R_g values for peak 1 ranged between 37.0 – 49.8 Å (average 46.2 ± 0.3 Å, Fig. 4-29C) and the D_{max} values between 108 - 186 Å (average 165 ± 4 Å, Fig. 4-29D, Table 4-6).

	Ensemble		Pool	
	Range	Average	Range	Average
Peak 1				
R_g (Å)	37 - 50	46.2 ± 0.3	27 - 51	37.7 ± 0.02
D_{max} (Å)	108 - 186	165 ± 4	83 - 213	135.7 ± 0.1
Peak 2				
R_g (Å)	34 - 58	45.7 ± 0.3	28 - 57	37.8 ± 0.03
D_{max} (Å)	100 - 218	166 ± 2	81 - 217	136 ± 0.2

Table 4-6. EOM R_g and D_{max} distribution for His-r*Pf*E2bc

The values for the pool ranged between a R_g of 27 – 51 Å (average 37.7 ± 0.02 Å) and a D_{max} of 83-210 (136 ± 0.1 Å). An extended flexible protein would have an R_g distribution close to that of the pool (in this case 24 Å), while a rigid protein R_g distribution would be within a few angstroms around the average value (Bernado *et al.*, 2007). The R_g distribution for His-r*Pf*E2bc was ~13 Å, thus the protein is flexible to some degree but not completely unfolded. A similar conclusion can be made from the D_{max} distribution. If His-r*Pf*E2bc were an extended unfolded protein, the variability in the D_{max} range would be that of the pool (127 Å). The D_{max} range for His-r*Pf*E2bc was ~80 Å, thus the protein is not completely unfolded but is flexible.

For His-r*Pf*E2bc peak 2, EOM with trimer symmetry yielded a slightly poorer fit to the scattering data (χ^2 = 1.04) than a simulation with monomer symmetry (χ^2 = 1.01, Fig. 4-30A & B). However, visual inspection of the curves shows the fit at low angles was better when modelling with trimer symmetry. The R_g and the D_{max} values ranged between 34 – 58 Å (average 45.7 ± 0.3 Å) and 100 – 218 Å (average 166 ± 2), respectively (Fig. 4-30C & D). The variation in the pool R_g and D_{max} distribution was 29 Å and 136 Å, respectively. The R_g distribution for His-r*Pf*E2bc peak 2 in this case was 24 Å, which makes it more

flexible than peak 1. The D_{max} distribution, 118 Å, was also indicative of a much more unstructured protein.

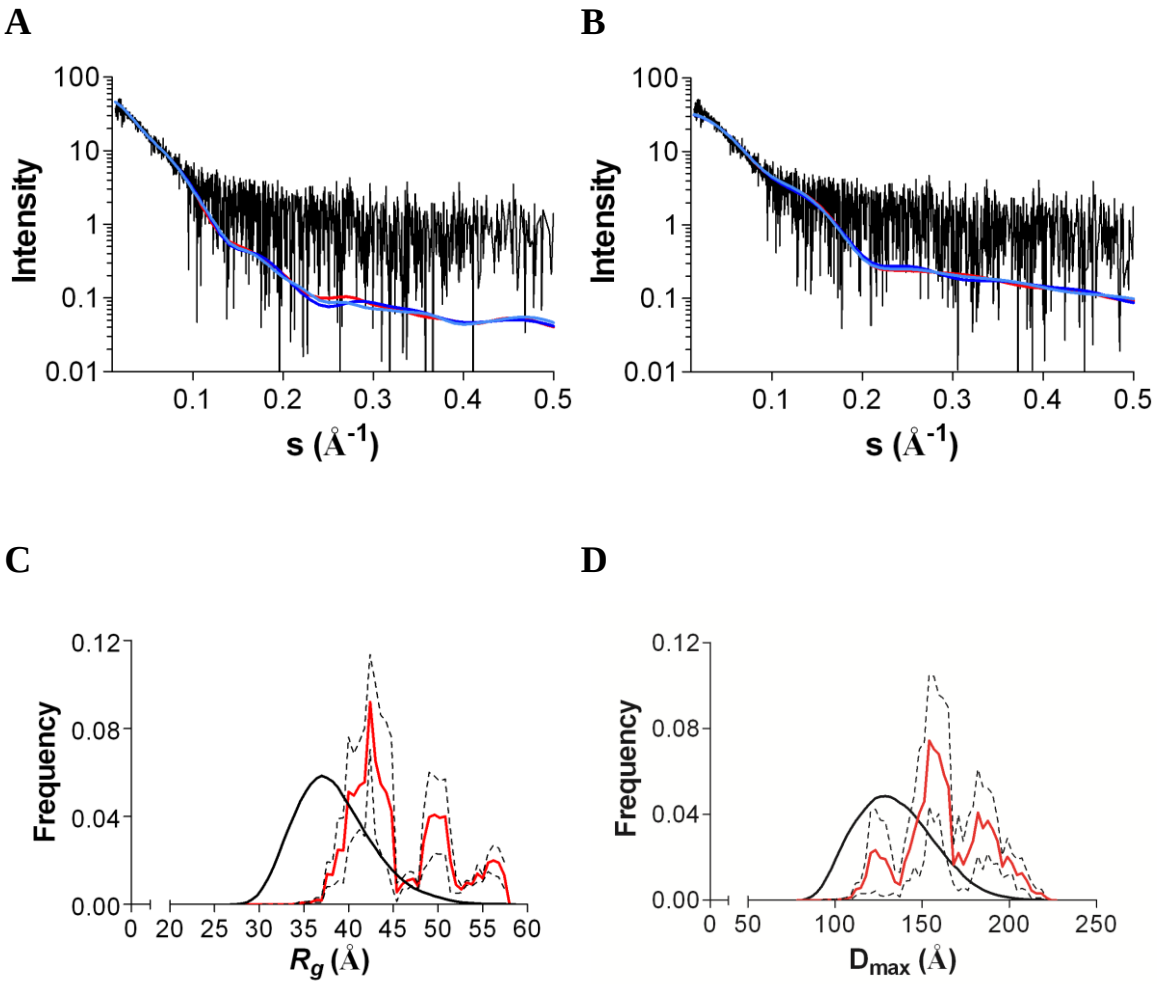


Figure 4-30. EOM of His-r*Pf*E2bc in peak 2
(A) Fit of ensemble curves from three EOM simulations using trimer symmetry (red, dark blue and light blue) to the scattering curve (black). (B) Fit of ensemble curve from three EOM simulations using monomer symmetry (red, dark blue and light blue) to the scattering curve (black). (C) R_g distributions of pool (black) and the average of three EOM simulations (red), with error shown as dashed line. (D) D_{max} distributions of pool (black) and the average of three EOM simulations (red), with error shown as dashed line.

EOM creates representative models for each of the ensembles. However, as the protein is flexible, any one model may not be the exact model for the protein. SAXS models are low resolution and because the EOM is carried out using high-resolution structures as templates, care must be taken not to over-interpret the resulting models. While the size distribution of an ensemble stays similar in repeat EOM simulations, the selected combinations of conformations can vary (Bernado *et al.*, 2007). However, the 10

representative models generated for peak 1 His-r*Pf*E2bc ensembles were of just two types (Fig. 4-31A&B).

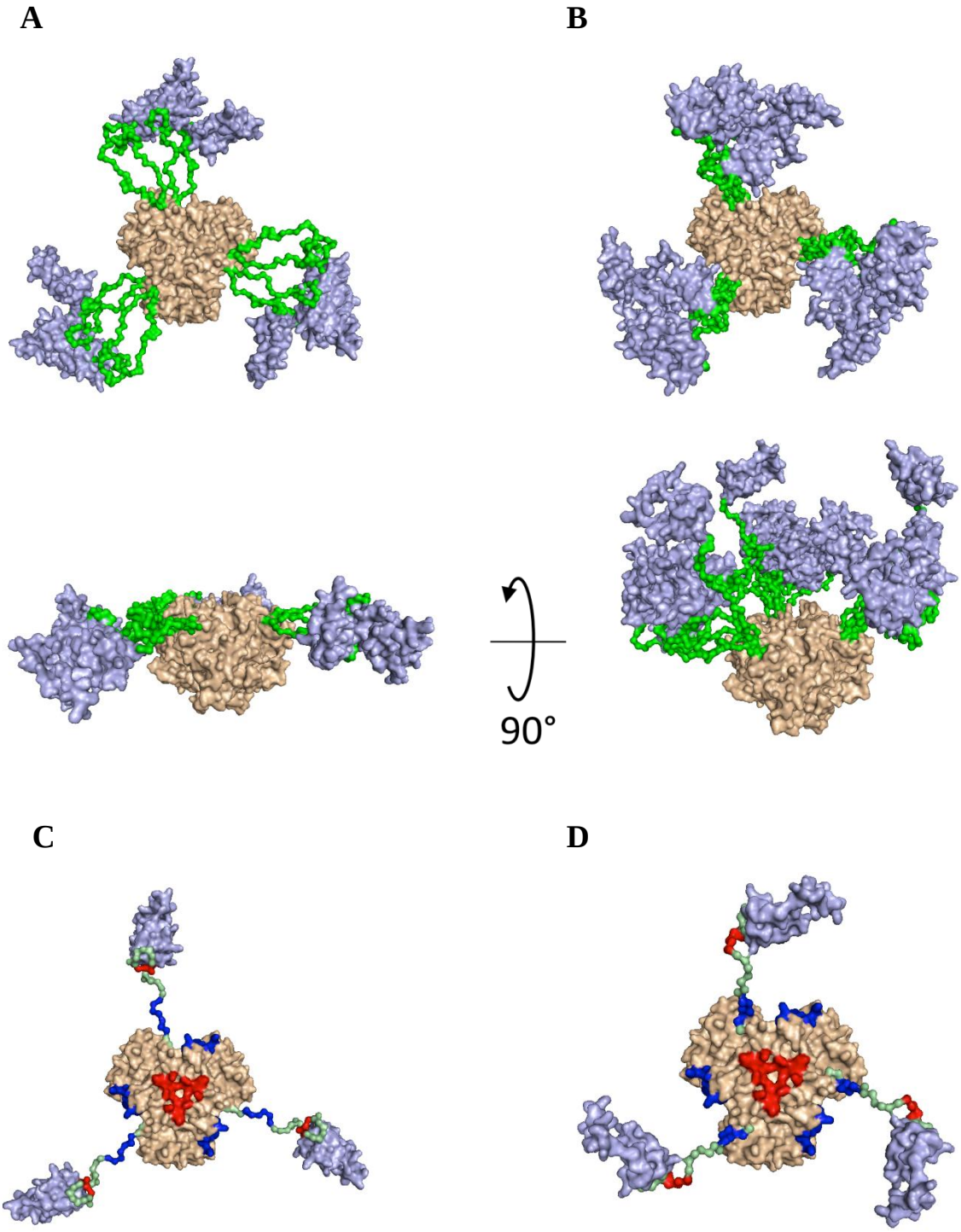


Figure 4-31. Representative models from EOM of peak 1 His-r*Pf*E2bc
(A) Group 1 contained 4 models, with an R_g range of 41.7 – 47.3 Å and a D_{max} range of 145 - 178 Å. (B) Group 2 comprised 6 models with an R_g range of 42.6 – 54.6 Å and a D_{max} range of 119 –

156 Å. (C) Position of residues possibly involved in intra-trimer interaction in group 1 models and (D) group 2 models. The blue residues in the linker (light green) could interact with the blue residues on the CD (light orange). Likewise for the red residues. The SBD is shown in light blue.

In one group of models the linker and the SBD point outwards on the x/y-axis (Fig. 4-31A) and in the other they point upwards on the z-axis (Fig. 4-31B). A similar grouping was seen with the DAMMIF models (Fig. 4-25 and 4-27A).

The residues in the linker, which are thought to be involved in trimer formation, are shown in Fig 4-31C & D. When these residues were modelled into the His-r*Pf*E2bc homology model, the SBD was oriented in a manner whereby it could not interact with E1 or E3 (Fig. 4-9 and 4-10). As these residues are based on a rigid crystal structure, it could be that in solution these interactions do not occur, or are transient. The extended linker formation in these EOM models would allow movement of the SBD to interact with its subunits.

As peak 2 His-r*Pf*E2bc was much more flexible, the 10 models that were generated were unsurprisingly much more varied with the position and the degree of extension of the linker and SBD, thus the models could not be easily grouped. The R_g range of the models was very wide, 31.8 – 57.8 Å, as was the D_{max} range, 106 – 203 Å. Because the outcome of every analysis on His-r*Pf*E2bc in peak 2 was more ambiguous than the protein in peak 1, no further analyses were done on this.

4.1.3.6 Hydrodynamic parameters of His-r*Pf*E2bc

US-SOMO (Rai *et al.*, 2005; Brookes *et al.*, 2010a; Brookes *et al.*, 2010a) was used to calculate the sedimentation coefficient of the I-TASSER homology models and the EOM models. The I-TASSER model for His-r*Pf*E2bc would have $s^0_{20,w}$ = 6.9 S, which is much larger than the experimental value for the protein in peak 1, which was 4.5 ± 0.3 S. This is due to the homology model being much more compact, thus it sediments faster. The computed $s^0_{20,w}$ for the EOM models (4.6 ± 0.1 S), correlates very well with the experimental value, giving further validation to the solution structure obtained for His-r*Pf*E2bc from the SAXS data.

4.2 Dihydrolipoamide dehydrogenase

4.2.1 Introduction

E3 is a highly conserved homodimer. This conformation is required for the formation of two active sites, which involve residues from both monomers (Mattevi *et al.*, 1991; Mattevi *et al.*, 1993a; Toyoda *et al.*, 1998; Brautigam *et al.*, 2005). A monomer of E3 has four domains; the FAD-binding domain, the NAD^+ binding domain, the central domain and the interface domain. Dimer formation involves six hydrogen bonds between residues in the interface domain and three in the FAD domain (Mattevi *et al.*, 1991).

The mature length apicoplast *Pf*E3 (*Pf*aE3), which is 63 kDa, is on average 12 kDa larger than E3 in other species. It is also larger than its mitochondrial counterpart (*Pf*mE3), which is 54 kDa. *Pf*aE3 contains the conserved residues required for FAD and NAD^+ binding, as well as all the active site residues (Cys165, Cys170, His646 and Glu651, McMillan, 2006). The structure of *P. falciparum* E3 is unknown, however, as the protein is highly conserved throughout various organisms, it is expected to be a structurally similar homodimer. As it was difficult to generate high quantities of His-r*Pf*aE3 for crystallisation, the protein was characterised using AUC and SAXS.

4.2.2 Aims

The aims for this part of the project were to:

- determine the oligomeric state of r*Pf*aE3
- obtain a structural model of r*Pf*aE3

4.2.3 Results

4.2.3.1 Homology modelling of *P. falciparum* E3

The I-TASSER server for protein structure and function prediction was used to obtain a homology model for the recombinant expression construct His-r*Pf*aE3 (http://zhanglab.ccmb.med.umich.edu/I-TASSER/, Zhang, 2008; Roy *et al.*, 2010, section 2.8.2). The top three templates used for the modelling were the *Pisum sativum* glycine cleavage system E3 (PDB entry 1DXL, Faure *et al.*, 2000), *Sinorhizobium meliloti* E3 (PDB entry 3URH, no associated publication) and *Sulfolobus soflataricus* E3 (PDB entry 3L8K, no associated publication). The best of the His-r*Pf*aE3 models had a C-score of -0.35 (Fig. 4-32A), while the other four models had scores ranging from -1.65 to -2.42. The His-tag (9 amino acids) was removed from the model for structural comparisons with E3 from other organisms. The best model had most similarity to the human E3 (PDB entry 1ZMC, Brautigam *et al.*, 2005) with rmsd=1.1 Å, *S. cerevisiae* E3 (PDB entry 1YEH, Toyoda *et al.*, 1998) rmsd=1.3 Å and *T. cruzi* E3 (PDB entry 2QAE, no associated publication) rmsd=1.4 Å. Similarity to *A. vinelandii* was the lowest with an rmsd of 2.2 Å (PDB entry 3LAD, Mattevi *et al.*, 1991). For comparison, the human and yeast E3, which are very similar, have an rmsd of about 0.8 Å (Brautigam *et al.*, 2005).

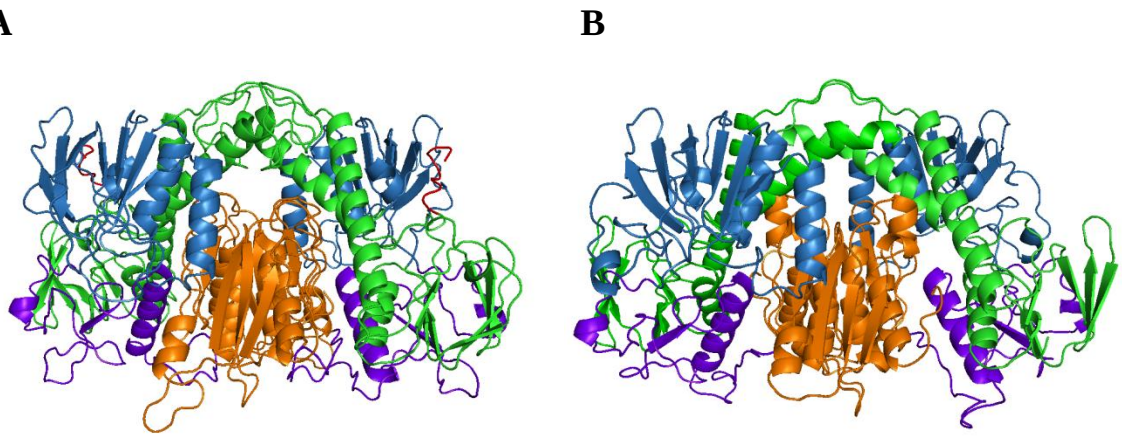


Figure 4-32. I-TASSER model of His-r*Pf*aE3 and *Pf*mE3
The FAD domain is in green (residues 1-191), the NAD^+ domain in blue (residues 193-334), the central domain in purple (residues 335-422) and the interface domain in orange (residues 423-557). (A) Homology model for His-r*Pf*aE3, the His-tag is in red. (B) *Pf*mE3 homology model. Residue numbers are those for the construct His-r*Pf*aE3.

An I-TASSER model was also generated for *Pf*mE3. The best model had a C-score of 0.43 (Fig. 4-32B), while the other four ranged from -0.89 to -1.82. In addition to the templates used for the His-r*Pf*aE3 model, the *P. falciparum* thioredoxin reductase was also used as a template (PDB entry 4B1B, Boumis *et al.*, 2012). E3 is part of the same protein family as thioredoxin reductase, thus some structural similarity would not be unexpected. Indeed, *Pf*mE3 aligns with *P. falciparum* thioredoxin reductase to an rmsd of 2.5 Å, making it more similar to the reductase than to *Pf*aE3 (rmsd = 3.2 Å). However, *Pf*mE3 is structurally most similar to *T. cruzi* E3 followed by yeast, human and *A.vinelandii* E3 (rmsd range 0.4-1.0 Å). *Pf*mE3 is more similar to the E3 from other organisms than it is to *Pf*aE3 (rmsd=1.3 Å). This could reflect the specificity of *Pf*aE3 for the apicoplast PDC, as the other E3s are shared between the KADCs in the mitochondrion.

The larger size of *Pf*aE3 (63 kDa versus 50 – 54 kDa for other E3s) is accounted for by six larger loops (numbered 1-6 in Fig. 4-33). Alignment of the *Pf*aE3 amino acid sequence with the I-TASSER template sequences shows gaps in the other organisms where the longer loops are present in *P. falciparum*. The loops do not contain any of the residues required for FAD or NAD^+ binding or catalytic activity, which are conserved amongst various organisms. As the residues involved in dimer formation are not very conserved, some of the loops could be involved in this. Loop 1 in the FAD domain is in close proximity to the second monomer, as is loop 6 (Fig. 4-32), thus residues in these could be involved in dimer formation. The other loops are at the periphery of the protein and do not seem to come into contact with other domains. Finally, the active site residues are in very similar positions as in the other organisms, as would be expected (Fig. 4-34).

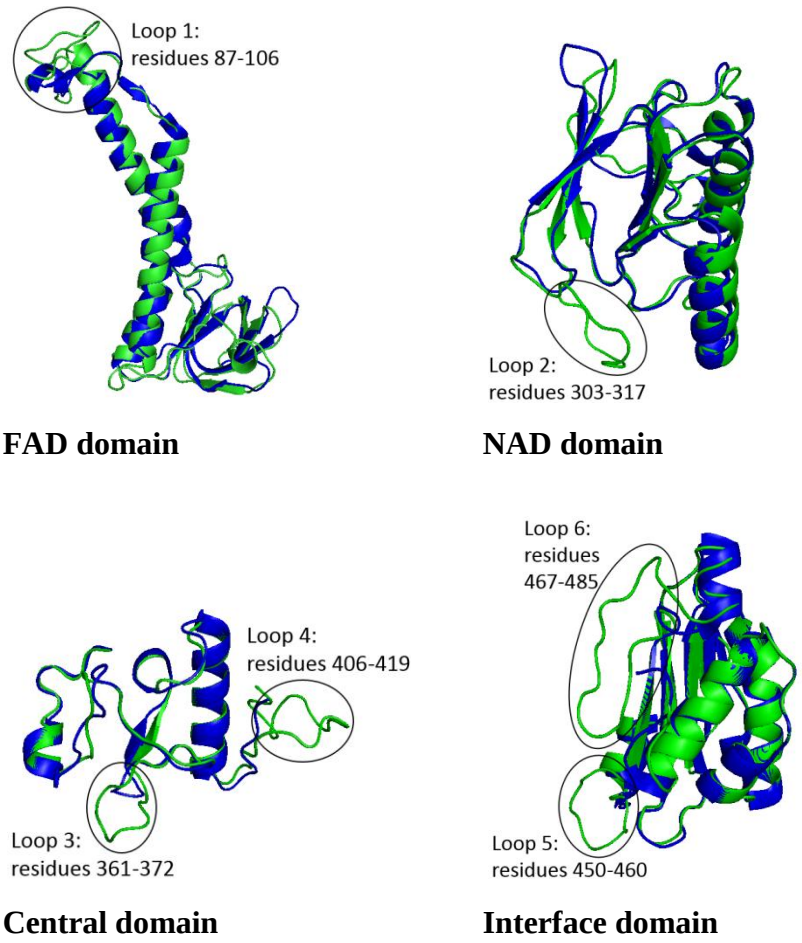


Figure 4-33. *Pf*aE3 homology model domains superimposed with corresponding human E3 domains
*Pf*aE3 is shown in green human E3 in blue. The larger loops present in *P. falciparum* are circled and the residue numbers shown.

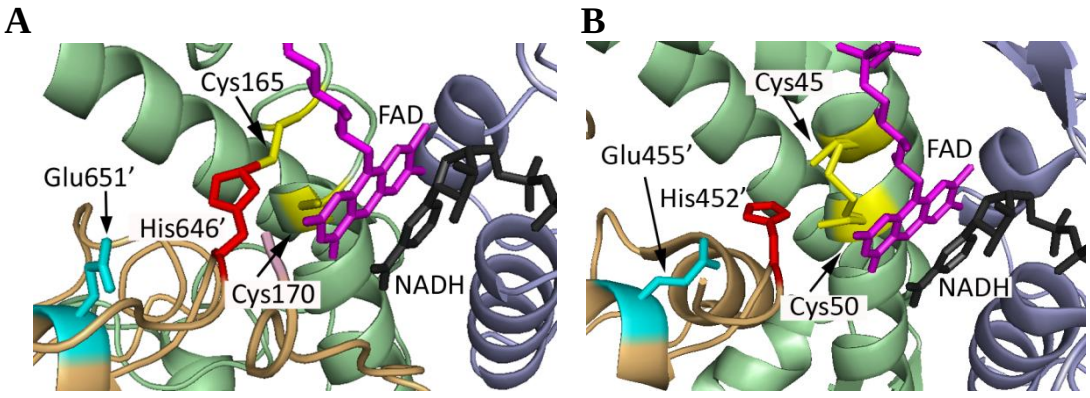


Figure 4-34. *Pf*aE3 active site
(A) *Pf*aE3 was aligned with human E3 to obtain the potential positions of FAD and NAD^+ in the *P. falciparum* I-TASSER model. The catalytic site comprises a disulphide (formed by Cys165 and Cys170), His646' from the second monomer, which is held in the correct conformation by Glu651', the isoalloxazine of FAD, and NAD^+. (B) Human E3 active site. Figure adapted from Brautigam *et al.*, (2005).

4.2.3.2 Stoichiometry of His-r*Pf*aE3

4.2.3.2.1 SV analysis of His-r*Pf*aE3

His-r*Pf*aE3 was analysed over a range of concentrations (3 – 23 μM) by SV. The reference buffer was PEBS250 (Table 2.1) and the density and viscosity values used for c(s) distribution analysis using SEDFIT (Schuck, 2000; Schuck *et al.*, 2002) were ρ = 1.01804 g/ml and η = 1.6354 cPoise, respectively. The partial specific volume of His-r*Pf*aE3 is calculated to be 0.7317 ml/g at 4ºC and 0.7385 ml/g at 20ºC (Table 2-8 and 2-9).

The samples consisted of a main species with $s_{20,w}$ = 6.0 – 6.1 S (68 – 94% of the sample) (Fig 4-35A and Table 4-7). A minor species (4 – 18%) of $s_{20,w}$ = 4.0 – 4.2 S was also present at all concentrations. Some samples (5, 6 and 16 μM) additionally contained a $s_{20,w}$ = 8.2 – 9.2 S species, albeit at a very low amounts (2 – 9%). No aggregation was seen, however, 1-13% of the sample was unaccounted for in the analysis (Table 4-7), which was most likely due to the degradation of His-r*Pf*aE3. The sedimentation coefficient at infinite dilution was determined to be $s^0_{20,w}$ = 6.1 ± 0.1 S (Fig. 4-35B).

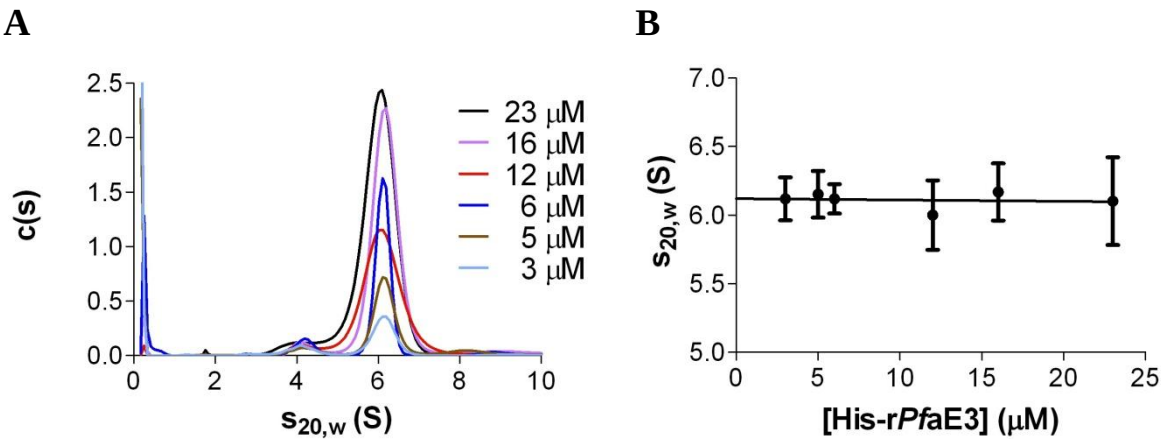


Figure 4-35. c(s) distribution analysis for a concentration range of His-r*Pf*aE3
(A) c(*s*) analysis of the SV data with SEDFIT (Schuck, 2000; Schuck *et al.*, 2002) was carried out for His-r*Pf*aE3 concentrations ranging from 3 to 23 μM. A main species with an average $s_{20,w}$ of 6.1 S and a minor ~4.1 S and ~8.8 S species can be observed. (B) A plot with His-r*Pf*aE3 concentration vs $s_{20,w}$ was generated to obtain the sedimentation coefficient at infinite dilution, which was $s^0_{20,w}$ = 6.1 ± 0.1 S. The experiment was only carried out once, thus, the error bars are the standard deviation shown as the square root of the central moment of the curve.

				% of total protein sample			
His-r*Pf*aE3 [μM]	$s_{20,w}$ (S)			~4.1 S	~ 6.1 S	~8.8 S	Sample loss
23	4.1	6.1	-	5	94	-	1
16	4.2	6.2	9.2	4	90	4	2
12	4.0	6.0	-	5	94	-	1
6	4.2	6.1	8.9	11	84	2	4
5	4.2	6.2	8.2	10	72	9	10
3	4.1	6.1	-	18	68	-	13

Table 4-7. Analysis of His-r*Pf*aE3 sample composition from c(*s*) analysis (Fig. 4-34).

4.2.3.2.2 SE analysis of His-r*Pf*aE3

For sedimentation equlibrium analysis, various concentrations (5-23 μM) of His-r*Pf*aE3 were centrifuged at 17,000 rpm (the speed was based on the predicted dimer mass of His-r*Pf*aE3) and 10 scans were taken 3 hours apart until the sample had reached equilibrium. The data aquired at each concentration were analysed using the non-interacting species analysis model in SEDPHAT (Vistica *et al.*, 2004, Fig. 4-36).

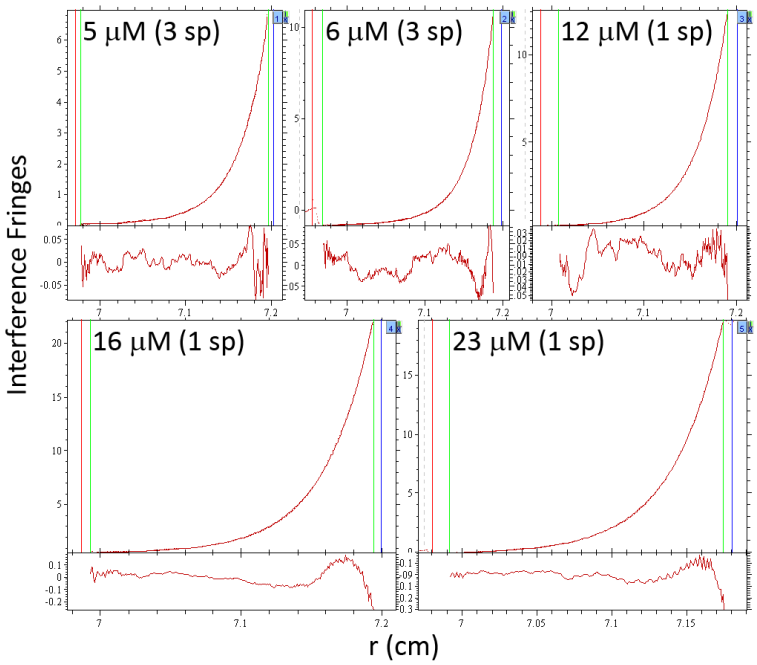


Figure 4-36. Analysis of His-r*Pf*aE3 SE data
Interference data (dots, not visible in plot as the data are so well fit by the model) was analysed with SEDPHAT (Vistica *et al.*, 2004) using the non-interacting single (1 sp) and three (3 sp) species model to obtain best fits (smooth line). The red vertical line indicates the cell meniscus, the blue vertical line indicates the cell base and the green vertical lines denote the range of data that was analysed. Residual plots are shown beneath each graph.

μM	MW (kDa)	% of sample
23	118	100
16	117	100
12	122	100
6	123	66
	62	11
	233	23
5	124	71
	60	16
	238	13

Table 4-8. Molecular mass of His-r*Pf*aE3 species at various concentrations as determined by SE.

The main species at the two lowest concentrations (5 and 6 μM) were 124 kDa and 123 kDa, respectively, which are slightly smaller than the estimated 128 kDa dimer size of His-r*Pf*aE3 (Table 4-8). The samples also contained a larger 233 - 238 kDa species, which could be a tetramer. In addition, both contained a small amount of a 60 kDa or 62 kDa species. This would be the size of a slightly truncated monomer (monomer size of His-r*Pf*aE3 is 64 kDa). This would be in agreement with the SV data, which show a minor species with $s_{20,w}$ ~4.1 S and the observation of a slightly smaller protein seen both on SDS-PAGE and on an anti-His-tag antibody western blot following gel filtration of His-r*Pf*aE3. This would suggest that the His-r*Pf*aE3 had been truncated at the C-terminus, thus preserving the N-terminal His-tag. Truncation at the C-terminus affects dimer stability. Deletion of just 14 amino acids from the C-terminus of *A. vinelandii* E3 has been reported to inactivate E3, decrease thermostability to almost that of a monomer and decrease FAD binding (Benen *et al.*, 1992). FAD binding is required for correct folding of E3 (Lindsay *et al.*, 2000), which, in turn, would affect dimer formation. A decrease from the estimated His-r*Pf*aE3 monomer size 64 kDa to 60-62 kDa would mean the loss of 15-34 amino acids, which would most likely prevent dimerisation of the protein.

The three higher concentration samples (12, 16 and 23 μM) contained one species which was 122 kDa, 117 kDa and 118 kDa, respectively (Table 4-9). These molecular masses are lower than the anticipated dimer size (128 kDa), suggesting degradation of His-r*Pf*aE3, however, possibly from the N-terminus rather than the C-terminus as above since the

protein has clearly dimerised. As the SE experiment required centrifuging the protein for several days, this could contribute to the degradation of the protein.

4.2.3.3 Solution structure of His-r*Pf*aE3

4.2.3.3.1 Small-angle scattering of His-r*Pf*aE3

SAXS measurements were carried out at the EMBL/DESY beamline X33. His-r*Pf*aE3 was purified two days before measurements and homogeneity was determined by SV analysis before and after measurements. A day before SAXS, His-r*Pf*aE3 was homogenous with a single 6.3 S species (Fig. 4-37). Two days after SAXS, some of the protein had degraded resulting in a small amount (5%) of a 3.9 S species in addition to the main 6.1 S species. Protein activity was also confirmed on the day of purification.

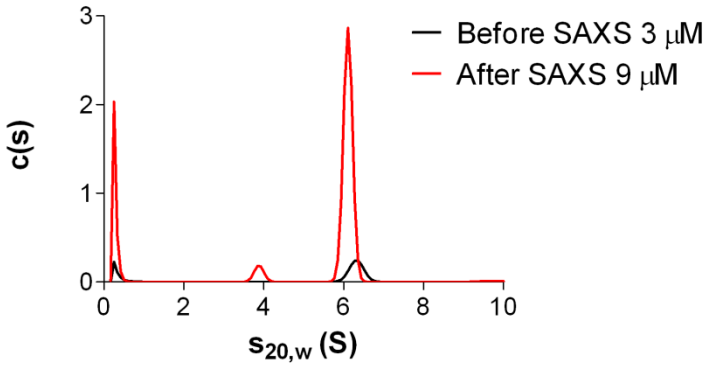


Figure 4-37. SV analysis of His-r*Pf*aE3 before and after SAXS
His-r*Pf*aE3 was homogenous before SAXS but had degraded a little after measurements. The peak seen near 0 S is due to buffer.

The final concentrations of His-r*Pf*aE3 used for SAXS measurements were 9 μM, 8 μM, 6 μM, 5 μM and 4 μM. High angle data from the highest concentration and low angle data from the lowest concentration were merged and the resulting scattering curve was used for all subsequent analyses (Fig. 4-38A). The linear Guinier region confirmed that there was no aggregation in the sample (Fig. 4-38B) and the radius of gyration, R_g, was determined to be 39.5 ± 0.3 Å.

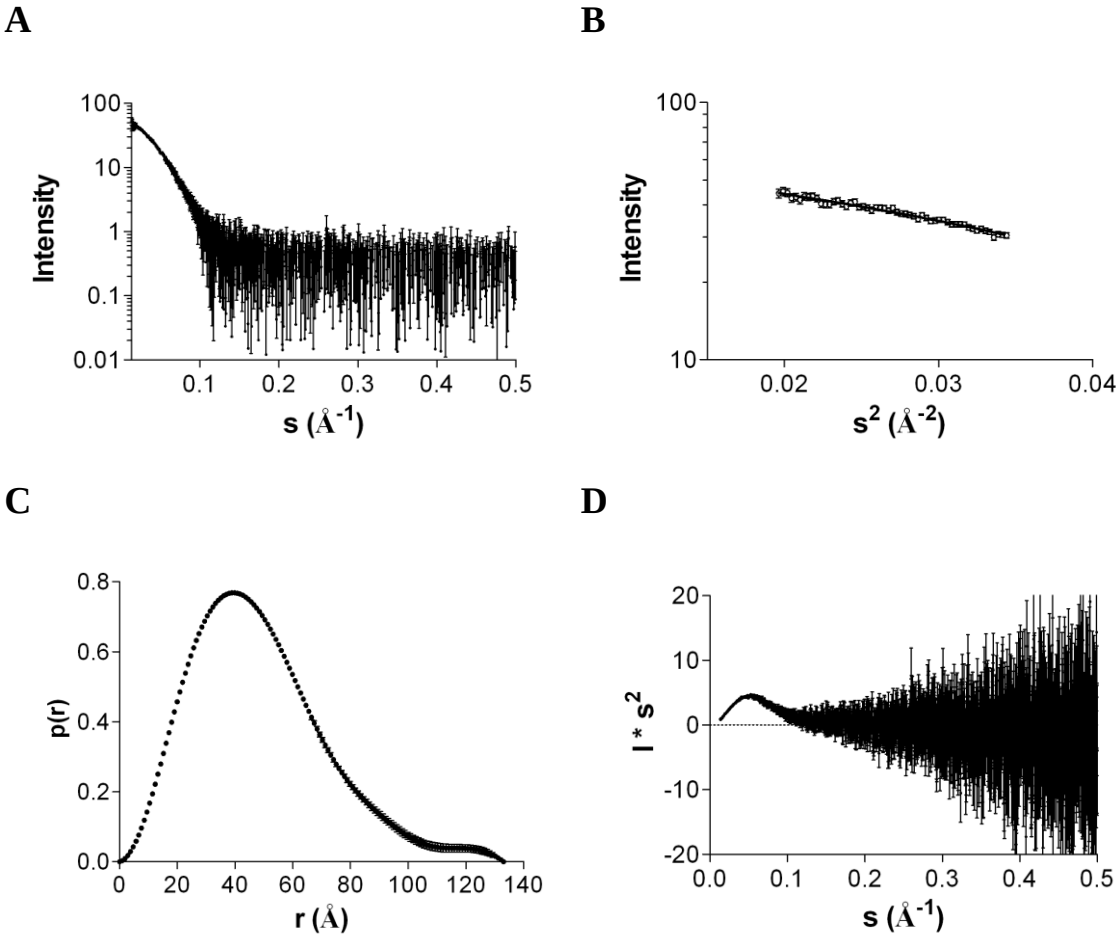


Figure 4-38. His-r*Pf*aE3 SAXS data analysis
(A) Experimental SAXS curve. (B) Guinier region. The R_g was 39.5 ± 0.3 Å and the *I*(0) was 53.1 ± 0.7. (C) *p(r)* distribution function.The D_{max} of His-r*Pf*aE3 was 133 Å. (D) Kratky plot shows the protein is folded.

GNOM (Svergun, 1992) was used to determine the distance distribution function, *p(r)*, from which the maximum dimension of the particle, D_{max}, was determined to be 133 Å (Fig. 4-38C). The R_g estimated from GNOM analysis, 37.8 ± 0.04 Å was similar to that determined from the Guinier region. The bell shape of the *p(r)* curve has a short tail, indicative of a globular particle which is slightly elongated. The Kratky plot also shows a bell-shaped curve, which means His-r*Pf*aE3 is folded (Fig. 4-38D).

The molecular mass of His-r*Pf*aE3 estimated by extrapolating the scattering intensity to zero angle, *I*(0) was 54 kDa. This would suggest that the sample was a monomer, however, the AUC data confirmed that His-r*Pf*aE3 was largely dimeric at concentrations used. The estimation of molecular mass from *I(0)* is not always reliable. Another method of determining the mass of a particle from SAXS data is by dividing the excluded volume of

the particle by 1.5 - 2. Determination of the excluded volume directly from scatting data is difficult when the data at high angles are very noisy, as in this case. The programme DATPOROD (Petoukhov *et al.*, 2012) can be used to determine the POROD volume from the *p(r)* distribution curve, which should improve accuracy. The POROD volume obtained was 188 nm^3, which would put the molecular mass estimation of His-r*Pf*aE3 between 94 - 125 kDa, the average being 110 kDa. This would be closer to the dimer size of 128 kDa.

4.2.3.3.2 *Ab initio* modelling of His-r*Pf*aE3

Ab initio modelling using P2 symmetry was carried out using DAMMIF (Franke and Svergun, 2009) on the ATSAS online server (http://www.embl-hamburg.de/biosaxs/atsas-online/dammif.php, Petoukhov *et al.*, 2007, section 2.8.3) to obtain 20 *ab initio* models, which were clustered by DAMCLUST (Petoukhov *et al.*, 2012). The models clustered into three groups of which the largest contained 15 models. Closer inspection showed slightly different models had been clustered together, thus clustering was carried out again with the 15 models. These were divided into two groups, one with seven and the other with eight models. The representative and the averaged model for both groups are shown in Fig. 4-39A - D.

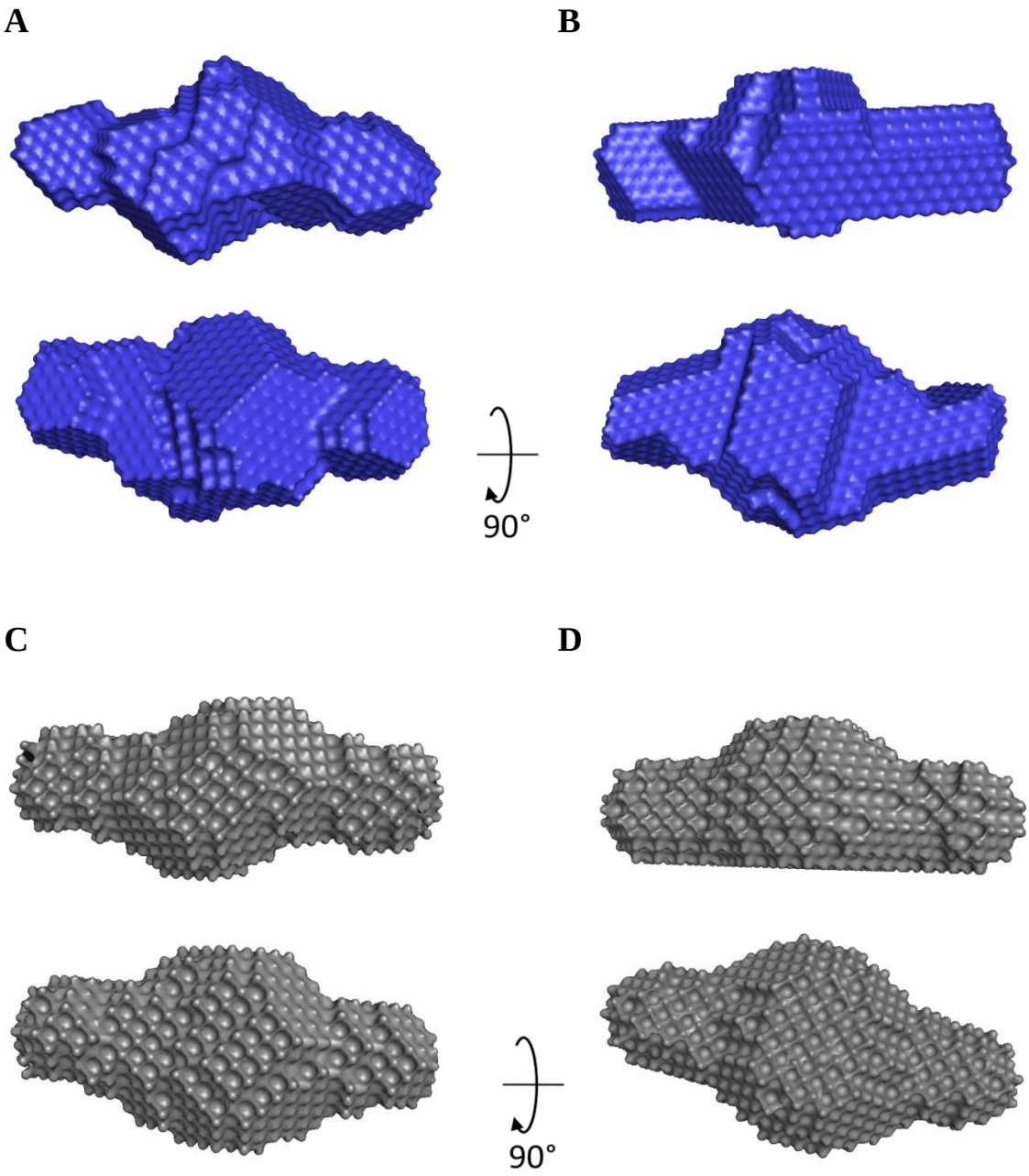


Figure 4-39. DAMMIF models of His-r*Pf*aE3
Representative model of (A) group 1 and (B) group 2. Average model of (C) group 1 and (D) group 2.

Both representative models had R_g = 37.8 Å. The group one model D_{max} was 146 Å (Fig. 4-39A), and group two D_{max} was 149 Å (Fig 4-39B). The molecular mass of the models was 128 kDa and 127 kDa, for group one and two, respectively. Both models fit the experimental scattering data well with Chi squared values of 0.79 (Fig. 4-40A & B).

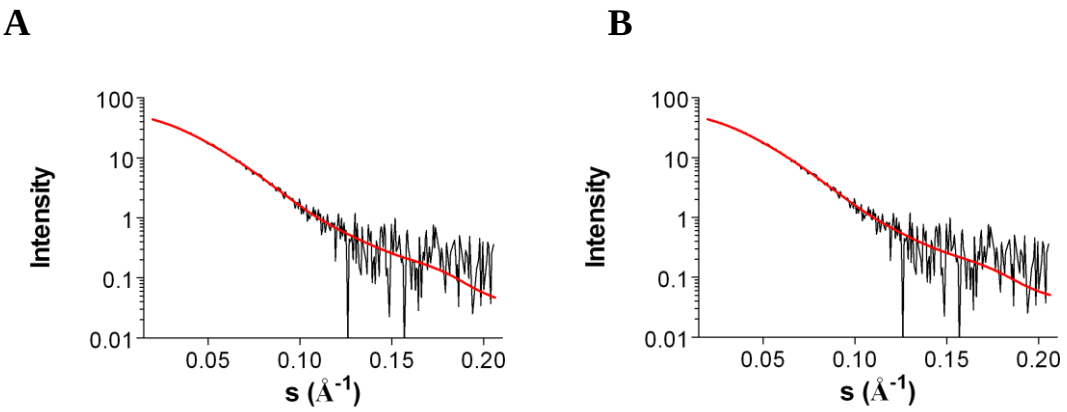


Figure 4-40. *Ab initio* model fits to scattering data for His-r*Pf*aE3
Representative model of (A) group 1 and (B) group 2. Scattering curve (black), *ab initio* model fit (red).

The averaged *ab initio* models were superimposed with the I-TASSER His-r*Pf*aE3 dimer model using SUPCOMB (Kozin and Svergun, 2001, Fig. 4-41). The normalised spatial discrepancy (NSD) values (section 2.8.4) for both superimpositions were 0.95. For comparison, NSD values close to one indicate that the two models are highly similar.

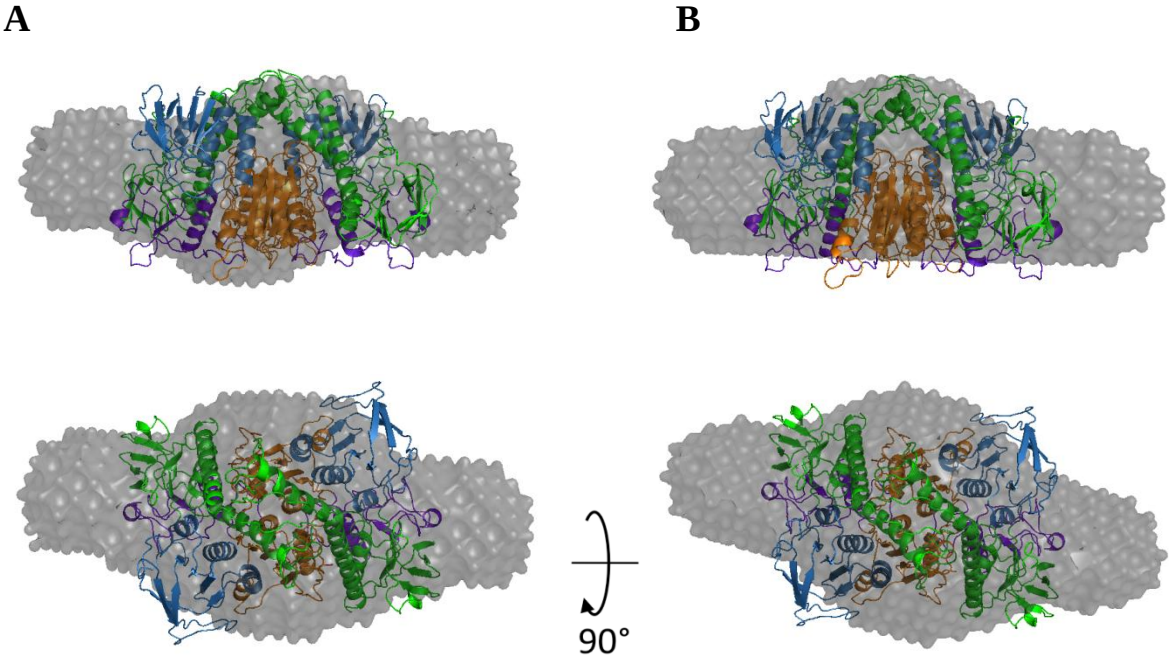


Figure 4-41. Superposition of I-TASSER homology model with His-r*Pf*aE3 *ab initio* model
Average model of (A) group 1 and (B) group 2. FAD domain (green), NAD domain (blue), central domain (purple), interface domain (orange).

The His-r*Pf*aE3 solution structure is more elongated than the I-TASSER model. The D_{max} of the homology model dimer was 110 Å, while the maximum dimensions of the averaged solution envelopes were 139 - 144 Å. The NAD^+ domains (blue) are partially outside the solution structure, while the homology model does not extend into the space neighbouring the N-terminus of the FAD (green) and central (purple) domains (Fig. 4-41).

There are a few flexible regions in His-r*Pf*aE3 (Fig. 4-42), as predicted by IUpred (Dosztanyi *et al.*, 2005, http://iupred.enzim.hu/). The N-terminus of His-r*Pf*aE3 contains a 14-amino acid flexible peptide including the His-tag, which could extend into the adjacent empty space in the DAMMIF model (Fig. 4-43, green). A loop in the NAD^+ domain extends outside the *ab initio* model (Fig. 4-43, blue). This is one of the longer loops in *Pf*aE3 (loop 2, Fig. 4-33) and is also slightly flexible according to IUpred. In the central domain, another longer loop extends outside the DAMMIF model (Fig. 4-43, cyan). The other extended loops in *Pf*aE3 (loops 1 and 4-6, Fig. 4-33) and a short flexible region (Fig. 4-43, yellow) cluster adjacent to another unoccupied region of the solution structure. It is likely that these loops and flexible regions are not well modelled by the rigid *ab initio* approach, manifesting themselves as extended lobes in the model in order to best fit the data. Instead, these regions should be explicitly modelled as flexible entities.

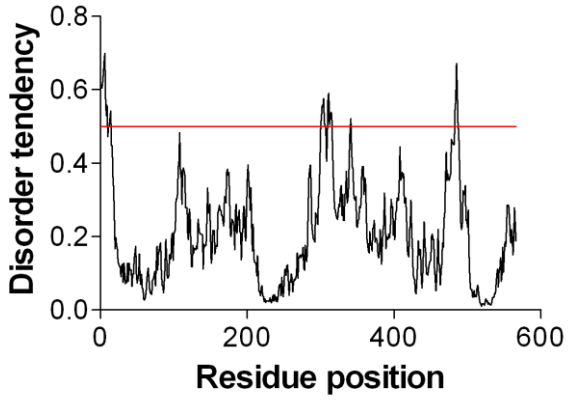


Figure 4-42. IUpred prediction of flexible regions in His-r*Pf*aE3
Values above 0.5 indicate flexibility. The flexible regions are residues 1-14, residues 300-315 in the NAD domain, and residues 483-487 in the interface domain (numbering according to His-r*Pf*aE3 construct).

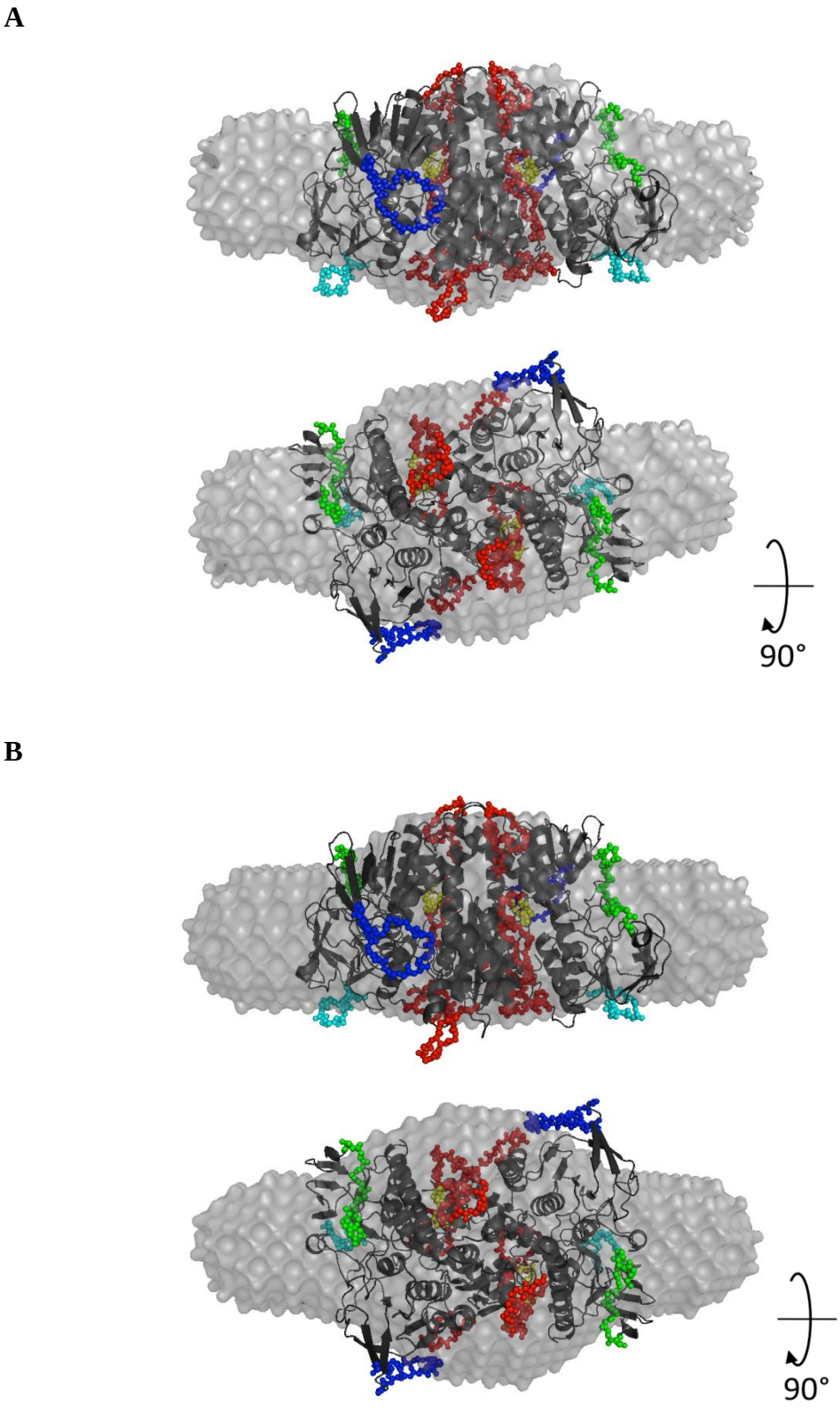


Figure 4-43. Flexible and extended loops in relation to the *ab initio* model of His-r*Pf*aE3
Average model of (A) group 1 and (B) group 2. His-r*Pf*aE3 N-terminus flexible region and His-tag is in green, *Pf*aE3 extended loop 2 in the NAD^+ domain is in blue and the central domain extended loop 3 in cyan. Loops 1 and 4-6 in the FAD, central and interface domain, respectively, cluster together (red). A shord flexible region is shown in yellow.

EOM (Bernado *et al.*, 2007; Petoukhov *et al.*, 2012) was carried out to assess whether the N-terminal flexible region with the His-tag could extend outwards to account for the empty space in the DAMMIF model. When changing the orientation of the His-r*Pf*E3 homology model in the DAMMIF model, the His-tag does seem to account for some of the space adjacent to it (Fig. 4-44). US-SOMO DMD (Dokholyan *et al.*, 1998; Ding and Dokholyan, 2006) simulations resulted in similar movement of the flexible N-terminal region.

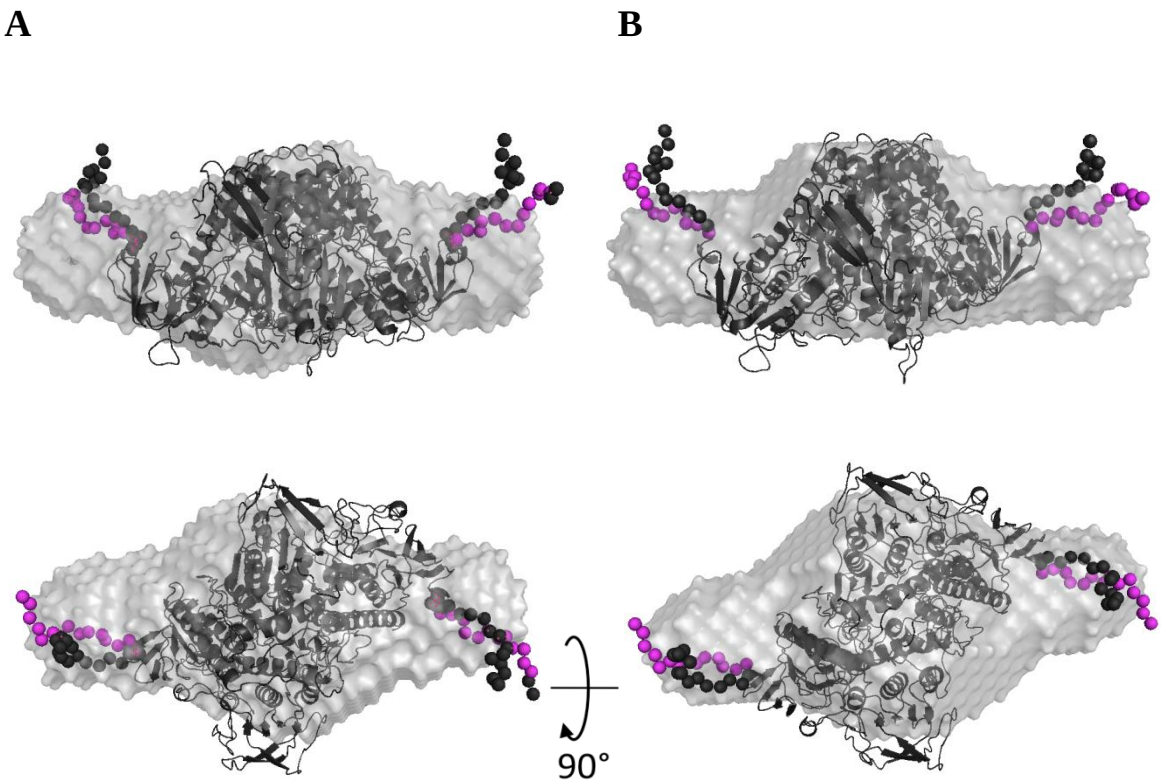


Figure 4-44. Modelling the flexible His-tag of His-r*Pf*aE3
Superposition of His-r*Pf*aE3 EOM model with the flexible 14-amino acid N-terminal His-tag with the DAMMIF average model for (A) group 1 and (B) group 2. EOM modelling was carried out twice, thus the His-tag is shown in grey and magenta.

4.2.3.4 Hydrodynamic modelling of His-r*Pf*aE3

Using the programme US-SOMO (Rai *et al.*, 2005; Brookes *et al.*, 2010a; Brookes *et al.*, 2010b) the sedimentation coefficients were calculated for the I-TASSER homology models and the *ab initio* models. The computed sedimentation coefficients also support the conclusion that the His-r*Pf*aE3 is more extended than the homology model. The I-TASSER dimer model has $s^0_{20,w}$ = 7.4 S, which is much larger than the experimental average of 6.1 S. This would suggest that the homology model is more compact that the real protein, as compact molecules sediment faster. The estimated $s^0_{20,w}$ value for the DAM model was 6.8 S, which is still higher than the experimental sedimentation coefficient for His-r*Pf*aE3. The computed sedimentation coefficients for the His-r*Pf*aE3 EOM models taking into account the flexible His-tag were 7.0 S. For the US-SOMO DMD models the sedimentation coefficients ranged from 7.0 to 7.3 S. Thus, the His-tag alone is not the source of the more elongated and less compact structure of His-r*Pf*aE3.

The calculated sedimentation coefficients for an I-TASSER monomer model and truncated monomers suggest that the monomer model itself is also more compact. The full length I-TASSER monomer would have $s^0_{20,w}$ = 4.6 S, which is higher than the average 4.1 S peak observed in the experimental data, attributed to a small amount of monomer. From the molecular mass of the species, 60-62 kDa (as determined by SE analysis, section 4.2.3.2.2), it can be deduced that this peak is the His-r*Pf*aE3 monomer, which is truncated by 15-34 amino acids. The computed $s^0_{20,w}$ for an I-TASSER monomer with 34 amino acids missing from the C-terminus was 4.4 S. This is still slightly higher than the experimental value. To obtain an $s^0_{20,w}$ of 4.2 S, 60 amino acids would need to be removed from the I-TASSER model. However, the molecular mass would then drop to 53 kDa, which is smaller than the experimental values. Thus, the extended and less compact structure of His-r*Pf*aE3 is not only due to how the dimer is assembled but is also influenced by the monomer structure of the protein. The truncation of the *Pf*aE3 C-terminus could be experimentally determined in the future by using mass spectrometry based methods to analyse the amino acid sequence of the C-terminus (Nakajima *et al*., 2012a, Nakajima *et al*., 2012b).

4.3 Summary

- Homology modelling using I-TASSER suggested that *Pf*E2 LD, SBD and CD structures have similar features to those from other organisms. Interestingly, differences in key residues in the *Pf*E2 SBD, which could be involved in E1 and E3 binding, were revealed. The I-TASSER homology *Pf*E2 CD was modelled into a trimer, which showed the potential inter-trimer interaction sites. The trimer models were also used to assess potential intra-trimer interactions required for large E2 core formation. This modelling suggested that interactions between the *Pf*E2 trimers could be weaker than those in other organisms.

- SV analyses of His-r*Pf*E2m showed no large core complex. The main species had a sedimentation coefficient of 6.1 ± 0.1 S, which was confirmed to be a trimer by SE analysis. The molecular mass (193-195 kDa) in the highest concentration samples was slightly lower than that calculated for His-r*Pf*E2m from its amino acid sequence (210 kDa), which suggests some degradation of the protein.

- SE analysis of His-r*Pf*E2m showed the presence of larger molecular mass species (1.5 – 3.7 MDa), which could be the E2 core complex. However, these species were not seen in SV analyses. Unfortunately, aggregates in the samples and thermodynamic non-ideality of the protein complicated the analyses of these species. Thus, the potential E2 core complex stoichiometry is inconclusive.

- His-r*Pf*E2bc only contained a trimer species in both SV and SE analysis. The protein had a sedimentation coefficient of 4.5 ± 0.3 S and a molecular mass of 104-108 kDa, which is consistent with the trimer size calculated from the amino acid sequence of His-r*Pf*E2bc (104 kDa).

- SAXS was carried out to obtain a solution structure of His-r*Pf*E2bc. Low concentration of the protein made analsis challenging, however, EOM models of His-r*Pf*E2bc fit the scattering data well at low angles (χ^2 = 1.03). The models also had computed sedimentation coefficients (4.6 ± 0.1 S) that correlated well with the experimental values for His-r*Pf*E2bc (4.5 ± 0.3 S).

- I-TASSER homology models of His-r*Pf*aE3 showed the larger size of *Pf*aE3 compared with E3 from other organisms to be due to longer loops between the conserved domains.

- His-r*Pf*aE3 is a dimer with a sedimentation coefficient of 6.1 ± 0.1 S and a molecular mass of 117-124 kDa, which is consistent with a slighty truncated protein as the molecular mass of His-r*Pf*aE3 calculated from the amino acid sequence is 128 kDa.

- The solution structure obtained for His-r*Pf*aE3 using SAXS was more elongated than the homology model based on the crystal strucutres of other organisms. The computed sedimentation coefficient for the homology model (7.4 S) is much larger than the experimental value for His-r*Pf*aE3 (6.1 ± 0.1 S). Surprisingly, the calculated sedimentation coefficient for the His-r*Pf*aE3 DAMMIF model was still high (6.8 S) regardless of the more elongated structure.

- Modelling the N-terminal flexible region including the His-tag did not significantly decrease the calculated sedimentation coefficient (7.0 S), thus the source of the more elongated shape is likely to be due to the long loops and flexible regions within the *Pf*E3 structure.

Chapter 5

P. falciparum dihydrolipoamide acetyltransferase and dihydrolipamide dehydrogenase gene deletion

5.1 Introduction

The *PfE2* gene (PlasmoDB: PF3D7_1020800, www.plasmodb.org) is 3192 bp, however it contains 10 predicted introns and complementary DNA analysis showed the full protein encoding sequence is 1923 bp (McMillan, 2006). The gene is located on chromosome 10, which is 1.7 Mb and predicted to encode 403 proteins of the overall ~5300 coding genes in the *P. falciparum* genome (Gardner *et al.*, 2002a; Gardner *et al.*, 2002b). *PfE2* is very A/T rich (78.6%) with the introns showing the characteristic higher A/T content than exons (87.5% and 72.9%, respectively) (McMillan, 2006). The 2001 bp *PfaE3* (PlasmoDB annotation PF3D7_0815900, www.plasmodb.org) does not contain introns (McMillan, 2006). It is located on chromosome 8, which is 1.3 Mb and encodes 312 proteins (Hall *et al.*, 2002). *PfaE3* has an A/T content of 74.7%.

A recent study of the rodent parasite, *P. yoelii*, showed that the PDC is essential only in the liver stages of the parasites. E1α and aE3 were essential for the progression of the parasites from the liver stage to the infective blood stage in mice (Pei *et al.*, 2010). Similarly, the type II fatty acid biosynthesis (FASII) pathway, for which PDC provides acetyl-CoA, was essential only for the liver stages (Yu *et al.*, 2008a). Nevertheless, reverse transcriptase-PCR and western blotting have shown that the genes are expressed in *P. falciparum* during the blood stages (Foth *et al.*, 2005; McMillan *et al.*, 2005). Even though the PDC may not be essential in the erythrocytic stages of *P. falciparum*, the proteins could still have a role in the growth and development of the parasite. The PDC enzymes could, for example, have a role in defence against oxidative stress via the reducing equivalents in the E2 and E3 enzymes (section 1.4.3.4).

5.2 Aims

The aims for this section of the project were to

- Generate *PfE2* and *PfaE3* deletion mutants
- Characterise the resulting Δ*PfE2* and Δ*PfaE3* parasite lines
    - Determine any effect of the deletion on the growth and development of blood stage parasites
    - Determine whether *Pf*E2 or *Pf*aE3 play a role in defence against oxidative stress

5.3 Results

5.3.1 Cloning of pCC-1 *PfE2* and *PfaE3* knockout constructs

pCC-1 and pCC-4 vectors were developed by Maier *et al*. (2006) for gene deletion by double cross-over recombination via the 5' and 3' flanks of the gene of interest using positive/negative drug selection. pCC-1 contains the *human dihydrofolate reductase* (*hdhfr*) gene and pCC-4 contains the *blasticidin-S-deaminase* (*bsd*) gene. These are used for positive selection of parasites containing the plasmid with WR99210 or Blasticidin-S-HCl, respectively. Both vectors contain the *S. cerevisiae cytosine deaminase* (*cyd*) gene as a bifunctional chimeric protein with *S. cerevisiae* uracil phosphoribosyl transferase, which forces integration of the plasmid into the gene loci by negative drug selection using 5-fluorocytosine (5-FC) (Maier *et al.*, 2006). The *cyd* gene converts 5-FC to 5-fluorouracil, which inhibits RNA synthesis and thymidylate synthase, resulting in death of parasites with episomal pCC plasmid. When double cross-over recombination takes place, the target gene is replaced with either the *hdhfr* or *bsd*, but the *cyd* gene is lost (Fig. 5-1A & B). Therefore, only parasites with plasmid integration should survive in the presence of 5-FC.

In this study, knockout constructs were made using the pCC-1 vector only (see section 2.5.2.4). As the *PfE2* gene sequence is extremely AT rich, there was difficulty finding optimum primer binding sites to produce the 5' and 3' end fragments for cloning. For the 5' end, a 506 bp fragment (nucleotides 685-1190) was generated, and for the 3' end, a 508 bp fragment (nucleotides 2119-2626) was produced. Although the resulting pCC1-Δ*PfE2* construct would not replace the whole *Pf*E2 gene, there would still be a deletion of 1941 bp from the centre of the 3192 bp gene. This would correspond to the loss of the protein from the second lipoyl domain onwards to the middle of the catalytic domain (Fig. 5-1A).

For the pCC1-Δ*PfaE3* construct, a 490 bp fragment (nucleotides 37-526) and a 545 bp fragment (nucleotides 1456-2001) were generated for the 5' and 3' flanks, respectively. Successful integration of the pCC1-Δ*PfaE3* construct into the *PfaE3* locus would delete the whole mature gene as the first 327 nucleotides are predicted to encode the signalling and transit peptides (Fig. 5-1B).

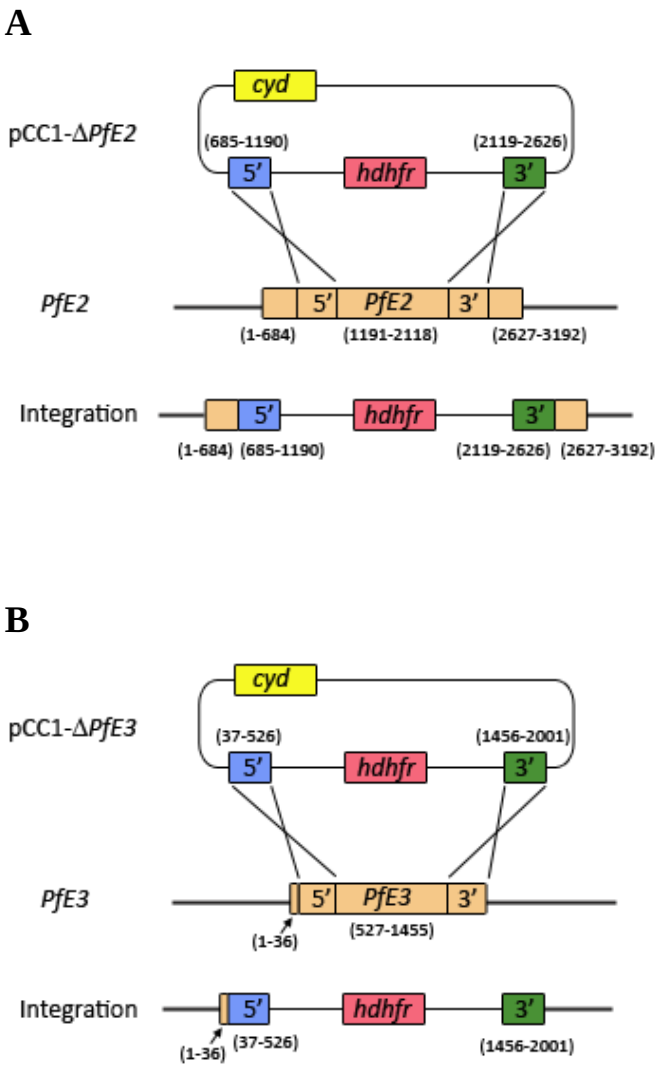


Figure 5-1. Schematic diagram for pCC-1 vector mediated double cross-over gene replacement of *PfE2* and *PfE3*.
The numbers in parenthesis are nucleotide sequence numbers for each gene portion. Diagrams are not to scale. (A) The pCC1-Δ*PfE2* plasmid contained ~500 bp flanks consisting of nucleotides 685-1190 at the 5' end and nucleotides 2119-2626 at the 3' end. If these flanks integrate into the corresponding sequences in the native gene, 1941 bp from the middle of the *Pf*E2 gene (nucleotides 1191-2118) are excised and replaced with the *hdhfr* selection marker. This deletion strategy leaves the start (nucleotides 1-684) and end (nucleotides 2627-3192) of *PfE2* in place, however, the majority of the gene is excised. (B) The pCC1-Δ*PfE3* plasmid contained ~500 bp flanks consisting of 5' nucleotides 37-526 and 3' nucleotides 1456-2001. When these flanks integrate into corresponding sequences in the native gene, nucleotides 527-1455 are replaced with the *hdhfr* selection marker. This deletes the whole *Pf*E3 gene as nucleotides 1-36 encode the transit peptide.

5.3.2 Generation of Δ*PfE2* and Δ*PfaE3 P. falciparum*

5.3.2.1 Positive/negative drug selection for *P. falciparum* with pCC1-Δ*PfE2* or pCC1-Δ*PfaE3* integration

The pCC1-Δ*PfE2* and pCC1-Δ*PfaE3* constructs were transfected in duplicate on two separate occasions into D10 and 3D7 parasite lines (section 2.9.8). The parasites were cultured with WR99210 for selection of transfectants and resistant parasites emerged after 3-7 weeks. Genomic DNA was isolated from the parasites (section 2.9.6 and 2.9.7) and Southern blotting (section 2.5.11) was carried out to confirm the presence of episomal pCC1-Δ*PfE2* or pCC1-Δ*PfaE3*. All the transfections were successful, apart from one transfection of 3D7 parasites with pCC1-Δ*PfE2*, from which no resistant parasites emerged.

Two of each construct/parasite line combination were chosen for further drug selection (section 2.9.9). Using 5-FC, negative drug selection pressure was applied to the cultures to force integration of the plasmids into their respective loci. 5-FC resistant parasites emerged in one pCC1-Δ*PfE2*/D10 parasite line (D10$^{pCC1\text{-}\Delta PfE2}$) and one pCC1-Δ*PfaE3*/3D7 parasite line (3D7$^{pCC1\text{-}\Delta PfaE3}$). DNA was harvested from these cultures and analysed for integration of plasmid by Southern blotting.

For genotyping of D10$^{pCC1\text{-}\Delta PfE2}$, gDNA was digested with *Hinc*II and *Bgl*I and initially probed for the *PfE2* 5' cloning fragment sequence. The expected DNA fragment sizes are shown in Fig. 5-2A. An 8.7 kb band potentially corresponding to full integration could be seen along with the 7.5 kb wild type (WT) D10 band (Fig. 5-2B). No band corresponding to the plasmid (5.5 kb) was present but two additional faint bands (~2.2 kb and ~1.4 kb) could be seen. As these did not correspond to any of the expected band sizes (Fig. 5-2A), the blot was stripped and the gDNA re-probed for the *cyd* gene. In the event of correct integration, this gene is lost and no bands should be seen with this probe. The 5.5 kb plasmid band was not seen, however, the smaller 0.75 kb band was present in the potential *Pf*E2 knockout line in addition to unidentified ~1.4 kb and 2.2 kb bands (Fig 5-2C), suggesting the presence of a population of parasites with incorrect integration of the plasmid. The *hdhfr* gene was also probed for, and only the 8.7 kb band corresponding to full integration was detected (Fig. 5-2D). In conclusion, at this stage these analyses

suggested that the D10$^{pCC1\text{-}\Delta PfE2}$ line did not contain an episomal copy of the pCC1-Δ*PfE2* plasmid and could be a mixture of parasites with both correct and incorrect integration of the plasmid. To obtain a homologous population of correctly integrated parasites, the line was subject to cloning by limited dilution (see section 5.3.2.2).

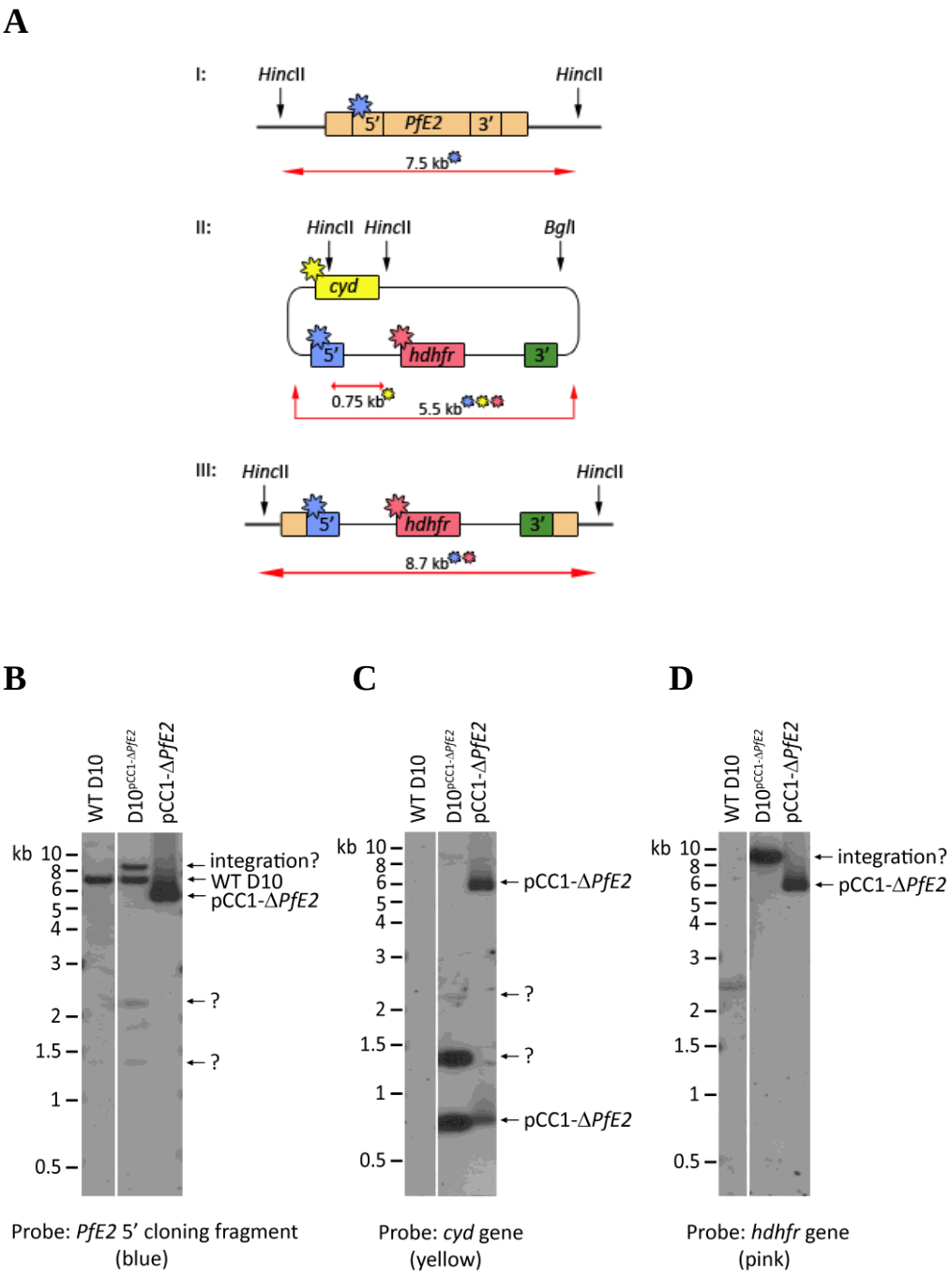


Figure 5-2. Southern blot analyses of pCC1-Δ*PfE2* integration into the *P. falciparum* D10 genome following positive/negative drug selection

(A) A schematic diagram (not to scale) of the expected DNA fragment sizes following digestion with *Hinc*II and *Bgl*I: (I) Endogenous gene: 7.5 kb fragment detectable with the *PfE2* 5' cloning fragment probe (blue star); (II) pCC1-Δ*PfE2* plasmid; 5.5 kb fragment detectable with all the probes, and 0.75 kb fragment detected only with the *cyd* probe (yellow star); (III) Integration: 8.7 kb fragment detectable by the *PfE2* 5' cloning fragment and *hdhfr* probes (blue and pink stars, respectively). (B) $D10^{pCC1\text{-}\Delta PfE2}$ gDNA (2.5 µg) digested with *Hinc*II and *Bgl*I. A potential full integration band (8.7 kb) and the WT D10 band (7.5 kb) were seen. No plasmid band (5.5 kb) was present. Two faint unidentified bands (~2.2 kb and ~1.4 kb) were detected. (C) Probed for the *cyd* gene, the 0.75 kb *cyd* fragment was detected in the $D10^{pCC1\text{-}\Delta PfE2}$ line. Also unidentified ~1.4 kb and 2.2 kb bands were present. (D) With a probe for the *hdhfr*, only a band corresponding to full integration (8.7 kb) could be seen.

The 3D7$^{pCC1\text{-}\Delta PfaE3}$ gDNA was digested with *Hinc*II and probed for the *PfaE3* 5' cloning fragment sequence. Only a 3.9 kb band corresponding to integration could be seen in the line, with no WT 3D7 or plasmid contamination (Fig. 5-3). Although the 3D7$^{pCC1\text{-}\Delta PfaE3}$ line seemed to have only a single population of integrated parasites, cloning by limiting dilution was also carried out to ensure a homogenous population.

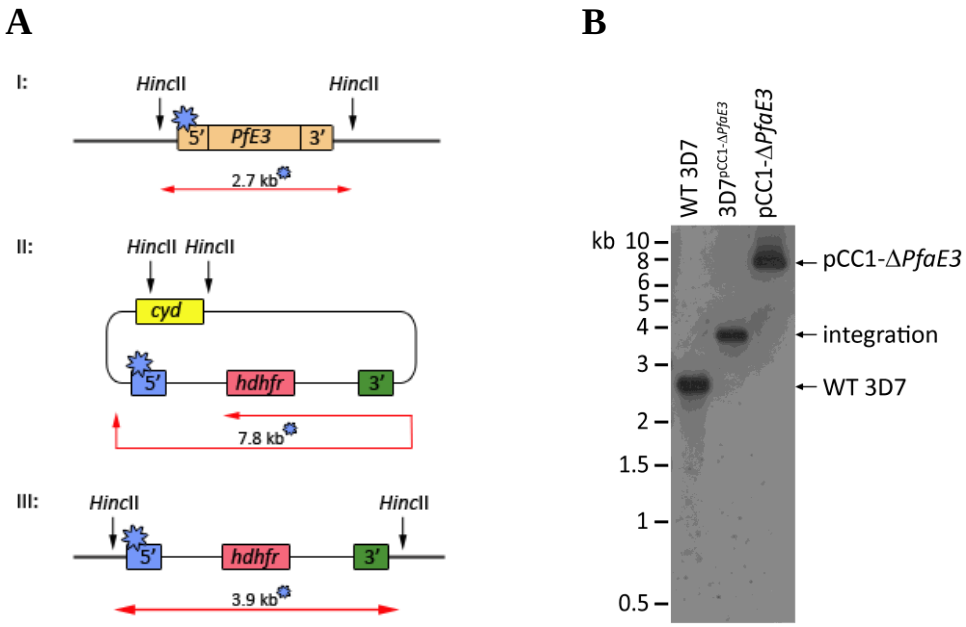


Figure 5-3. Southern blot analysis of pCC1-Δ*PfaE3* integration into *P. falciparum* 3D7 genome following positive/negative drug selection
(A) A schematic representation (not to scale) of the expected band sizes following digestion with *Hinc*II. The *PfaE3* 5' cloning fragment probe (blue star) was used to visualise DNA. (I) Digestion of the endogenous gene results in a 2.7 kb fragment. (II) Digestion of the pCC1-Δ*PfaE3* results in a 7.8 kb fragment. (III) Following integration pCC1-Δ*PfaE3* plasmid, digestion with *Hinc*II results in a 3.9 kb fragment. (B) 3D7$^{pCC1\text{-}\Delta PfaE3}$ gDNA (2.5 µg) digested with *Hinc*II probed for the *PfaE3* 5' cloning fragment sequence. Only a band corresponding to correct integration (3.9 kb) could be seen in the 3D7$^{pCC1\text{-}\Delta PfaE3}$ line. No bands corresponding to WT 3D7 (2.7 kb) or pCC1-Δ*PfaE3* plasmid (7.8 kb) were present in this line.

5.3.2.2 Cloning of integrated *P. falciparum* lines by limited dilution

To obtain homogenous populations of knockout parasites, cloning by limiting dilution was carried out (section 2.9.10). The D10$^{pCC1\text{-}\Delta PfE2}$ and 3D7$^{pCC1\text{-}\Delta PfaE3}$ lines were diluted to 0.25 parasites/well in culture medium containing both WR99210 and 5-FC into two 96-well mictrotitre plates. After 3-4 weeks, 17 D10$^{pCC1\text{-}\Delta PfE2}$ and 34 3D7$^{pCC1\text{-}\Delta PfaE3}$ clonal lines had emerged. These were then bulked to larger cultures, stored in liquid nitrogen and genomic DNA was extracted for genotypic analysis.

Genomic DNA of four clones of the potential *PfE2* knockout parasite line D10$^{pCC1\text{-}\Delta PfE2}$ was digested with *Hinc*II and *Bgl*I and probed for the *PfE2* 5' cloning fragment sequence. Three out of the four clones seemed to show the 8.7 kb band possibly corresponding to integration, however, the 7.5 kb WT band was still also present (Fig. 5-4). One of the clones had only the 7.5 kb WT band. In addition, the several unidentified bands were still seen casting doubt on whether the line was an actual *PfE2* knockout line as was first thought.

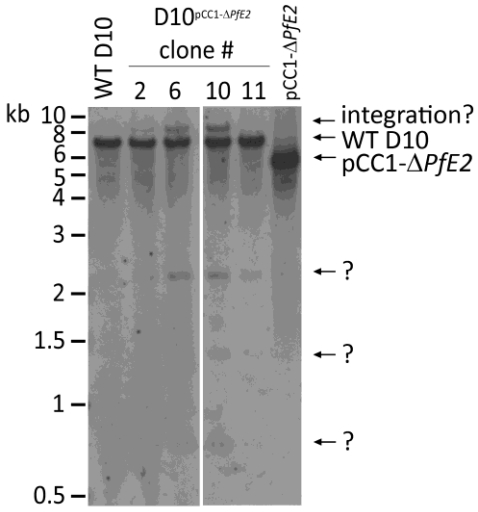


Figure 5-4. Southern blot analyses of potential clonal *PfE2* knockout *P. falciparum* lines
gDNA (2.5 µg) from four clones (numbered 2, 6, 10 and 11 in order of emergence following limiting dilution) was digested with *Hinc*II and *Bgl*I and probed for the *PfE2* 5' cloning fragment sequence. A possible intergration band (8.7 kb), the WT (7.5 kb) band and several unidentified bands were present in the potential knockout line. For expected fragment sizes see Fig. 5-2A.

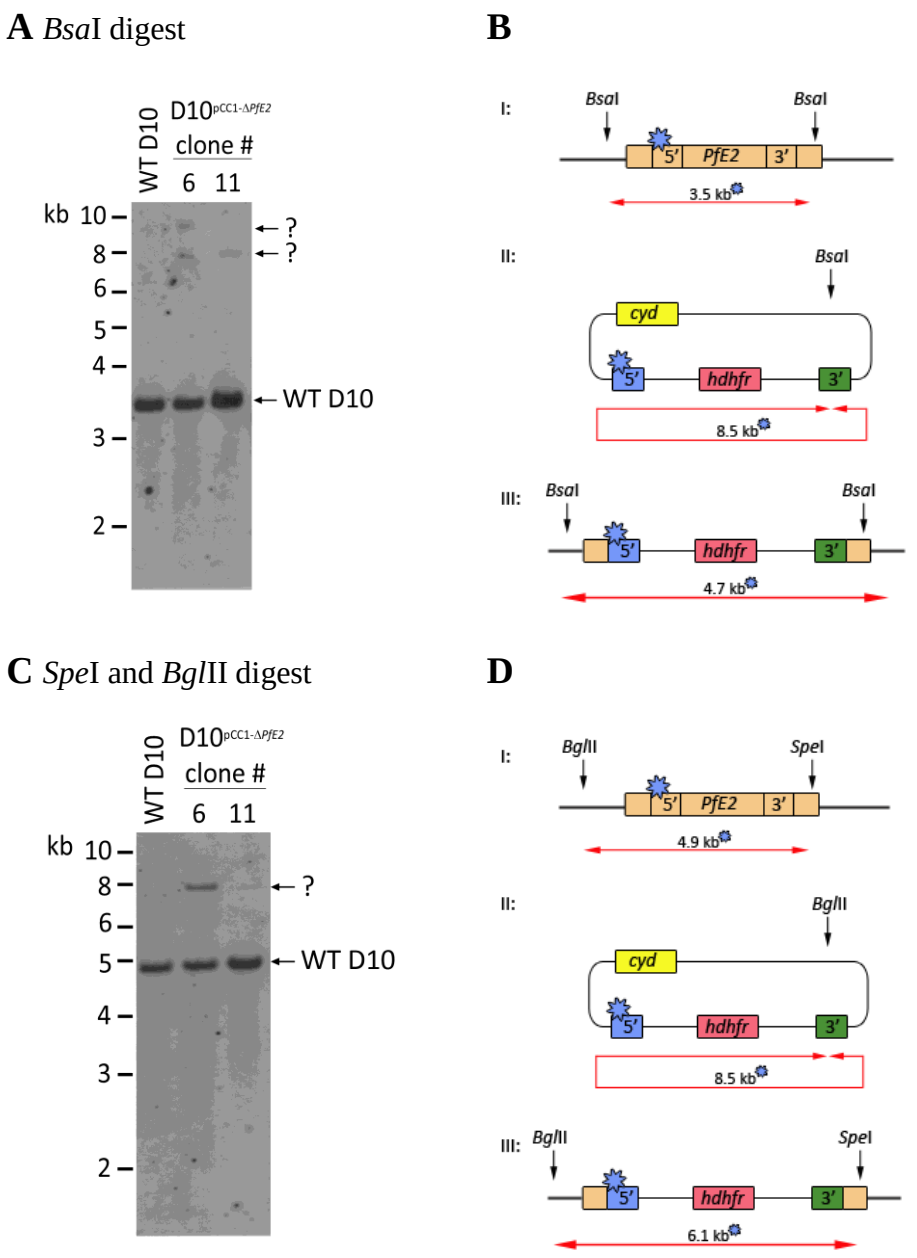


Figure 5-5. Additional diagnostic Southern blot analyses of potential clonal *PfE2* knockout *P. falciparum* lines

(A) A digest with *Bsa*I probed for the *PfE2* 5' cloning fragment sequence showed only the WT 3D7 band (3.5 kb) and two unidentified bands (~8 kb and ~9 kb). (B) A schematic diagram (not to scale) of the expected DNA fragment sizes following digestion with *Bsa*I. The *PfE2* 5' cloning fragment probe (blue star) was used to visualise DNA. (I) Endogenous gene: 3.5 kb fragment. (II) pCC1-Δ*PfE2*: 8.5 kb fragment (III) Integration: 4.7 kb fragment. (C) A digest with *Spe*I and *Bgl*II and probed for the *PfE2* 5' fragment sequence also showed no integration band (6.1 kb). Only WT D10 (4.9 kb) and an unidentified ~8 kb band were seen. (D) A schematic representation (not to scale) of the expected DNA fragment sizes following digestion with *Spe*I and *Bgl*II. The *PfE2* 5' fragment probe (blue star) was used to visualise DNA. (I) Endogenous gene: 4.9 kb fragment. (II) pCC1-Δ*PfE2*: 8.5 kb fragment. (III) Integration: 6.1 kb fragment.

To confirm the genotype, two additional digests were carried out on two of the clones. A digest with *Bsa*I showed only the WT D10 band (3.5 kb) and two unidentified bands (~8 kb and ~9 kb) (Fig. 5-5A). Correct integration would have produced a 4.7 kb band (Fig 5-5B). A digest with *Spe*I and *Bgl*II showed similar results; there was no integration band (6.1 kb), only a WT D10 (4.9 kb) and an unidentified ~8 kb band (Fig. 5-5D&E). Thus, the conclusion was made that the pCC1-Δ*PfE2* plasmid had not targeted the *PfE2* locus and no further work was carried out with the parasite line.

Five 3D7$^{pCC1\text{-}\Delta PfaE3}$ clonal lines were analysed by Southern blot (Fig. 5-6). The gDNA was digested with *Hinc*II and probed for the *PfaE3* 5' cloning fragment as before. All five lines showed the presence of only the band corresponding to integration (3.9 kb), confirming these to be homogenous populations of *PfaE3* knockout *P. falciparum*. 3D7$^{pCC1\text{-}\Delta PfaE3}$ clones number 16 and 32 were used for phenotypic studies.

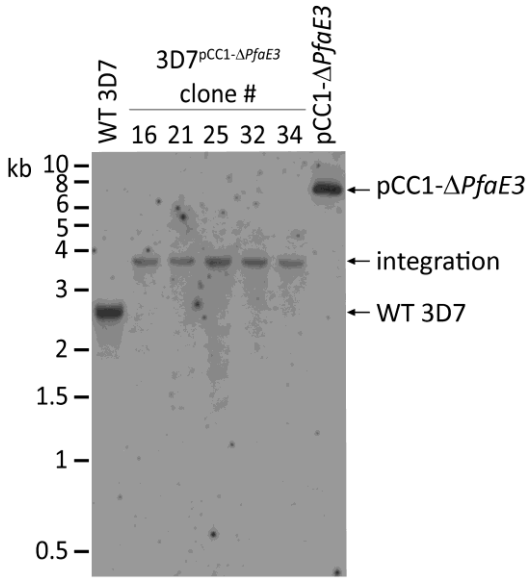


Figure 5-6. Southern blot analysis of 3D7$^{pCC1\text{-}\Delta PfaE3}$ clonal lines
gDNA (2.5 µg) from five clones (numbered 16, 21, 25, 32 and 34 in order of emergence following limiting dilution) was analysed and confirmed to be homogenous *PfaE3* knockout parasites as only the 3.9 kb integration fragment could be seen. The gDNA was digested with *Hinc*II and probed for the *PfaE3* 5' cloning fragment sequence. For the expected fragment sizes see Fig. 5-3A.

5.3.3 3D7$^{pCC1\text{-}\Delta PfaE3}$ growth assay

To investigate the effect of the deletion of *Pf*aE3 on parasite development, growth of the parasites was observed for 15 days (section 2.9.11). The 3D7$^{pCC1\text{-}\Delta PfaE3}$ mutant showed impaired growth in erythrocytic stage parasites compared with the wild type parental strain (Fig. 5-7). This is contrary to *P. yoelii*, in which deletion of *aE3* did not have an effect on parasite growth in the blood stages (Pei *et al.*, 2010). The initial multiplication rate (i.e. how much the parasitemia increased after the first 48 h cycle) was 8.6 ± 0.4 for the WT 3D7 compared with 6.4 ± 0.7 for 3D7$^{pCC1\text{-}\Delta PfaE3}$ clone 34. The parasitemia of WT 3D7 reached 12.2 ± 1.4 % on day 5, after which it steadily decreased until no parasites were observed in the culture on day 14. This suggests that the WT 3D7 reached its limiting parasitemia on day 5, after which parasite death started to occur. Such death of parasites was not observed in the 3D7$^{pCC1\text{-}\Delta PfaE3}$ clones, where the parasitemia fluctuated between 2.3 ± 0.5 % and 4.6 ± 0.2 % on the second day of each cycle throughout the experiment.

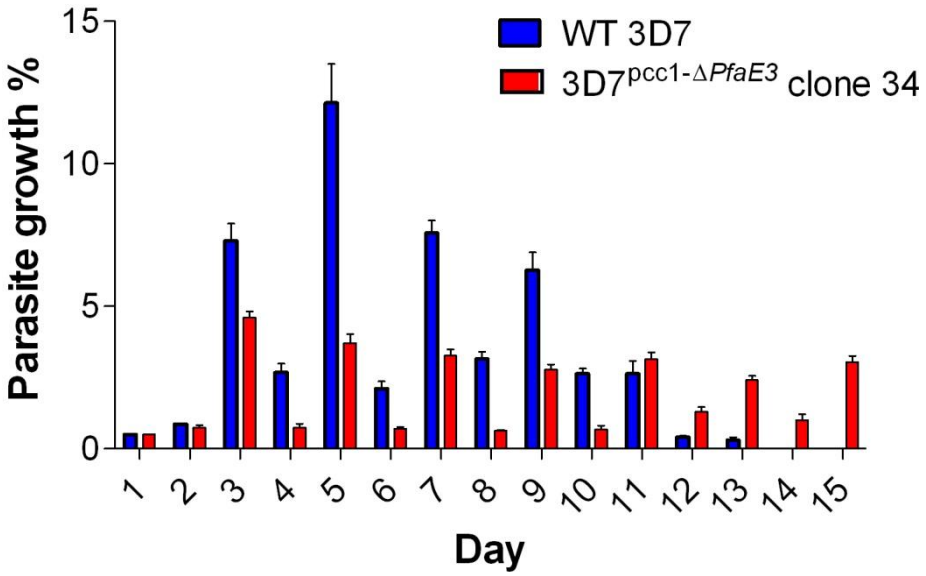


Figure 5-7. Growth of 3D7$^{pCC1\text{-}\Delta PfaE3}$
Synchronised ring stage cultures (in triplicate) were diluted to 0.5 % parasitemia (day 1) and every second day the cultures were diluted 1:5 with fresh erythrocytes. Blood smears of the cultures were made every 24 h. Parasitemia was determined by counting the number of infected erythrocytes per 1000 erythrocytes (section 2.9.2). The error bar denotes the standard error of the mean.

The cell cycle stage was also monitored (Fig. 5-8). As the cultures were synchronised on day 1 all the parasites were rings. On day 2 most parasites were in trophozoite stage. The WT 3D7 culture contained some schizonts (3 ± 1 %), while 3D7$^{pCC1\text{-}\Delta PfaE3}$ clone 34 only had a small proportion of schizonts (0.7 ± 0.7 %). On day 3 following invasion, more rings had already progressed to trophozoite stage (24 ± 1.4 %) in WT 3D7 than in 3D7$^{pCC1\text{-}\Delta PfaE3}$ clone 34 (4.3 ± 0.7 %). These differences became more obvious in the next three cycles (days 3-8) before the WT 3D7 culture began dying. This would imply that progression of the 3D7$^{pCC1\text{-}\Delta PfaE3}$ clone 34 through the cell cycle was slowed, thus resulting in the lower parasitemias observed during the growth experiment. Also the 3D7$^{pCC1\text{-}\Delta PfaE3}$ clone stayed highly synchronous over a lengthy period of 7 days, while the WT 3D7 parasites lost their synchrony after just two cycles (4 days).

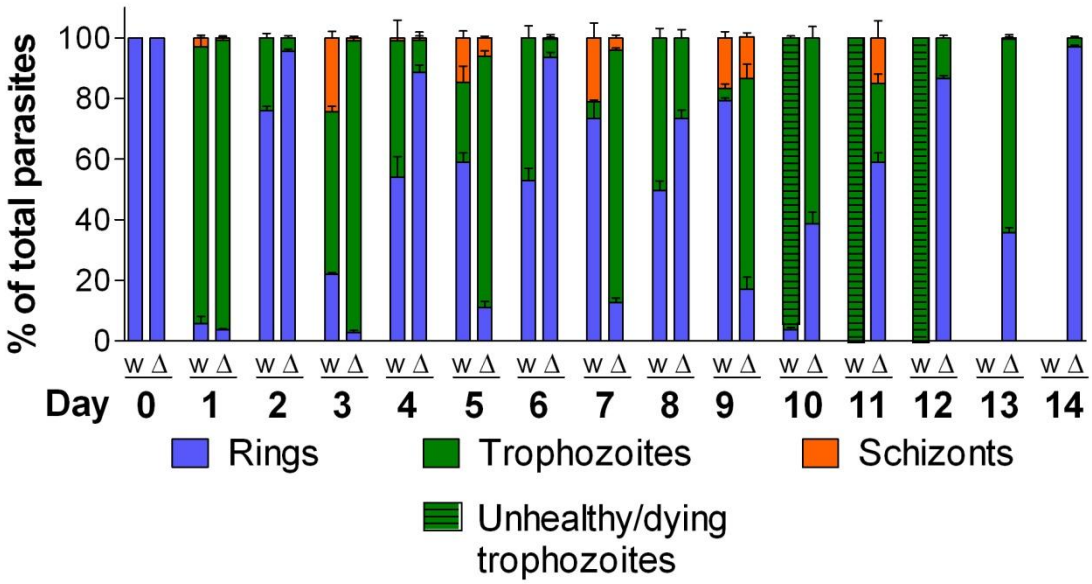


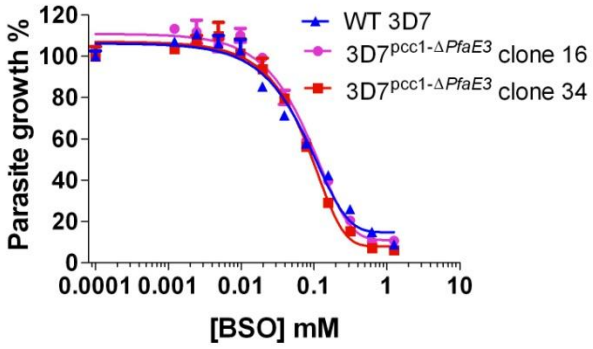
Figure 5-8. Cell cycle stage of 3D7$^{pCC1\text{-}\Delta PfaE3}$ compared with WT 3D7
The cycle stage of the parasite lines (in triplicate) was determined by counting 100 parasites and noting the number of rings, trophozoites and schizonts present. The error bar denotes the standard error of the mean. Trophozoites containing several large vacuoles or those that were extremely dense were classed as unhealthy/dying trophozoites.

5.3.4 3D7$^{pCC1\text{-}\Delta PfaE3}$ susceptibility to oxidative stressors and triclosan

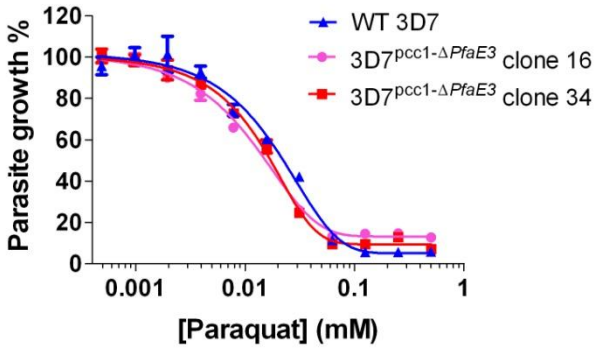
To begin investigation into the role *Pf*aE3 may play during the erythrocytic stages, susceptibility of 3D7$^{pCC1\text{-}\Delta PfaE3}$ to two different oxidative stressors and a drug thought to affect the FASII pathway was determined. L-buthionine sulfoximine (BSO) is a specific inhibitor of γ-glutamylcysteine synthase (γ-GCS), an enzyme in the glutathione synthesis pathway, which also affects *P. falciparum* growth (Griffith, 1982; Meierjohann *et al.*, 2002). Inhibiton of γ-GCS results in reduced levels of the important antioxidant glutathione, which is essential for blood stage parasite survival (Luersen *et al.*, 2000; Patzewitz *et al.*, 2012). Paraquat is a commonly used herbicide, which produces superoxide anions and oxidises cellular NADPH levels and causes lipid peroxidation (Suntres, 2002). Paraquat is also used experimentally, for example, to study responses to oxidative stress in *P. falciparum* (Komaki-Yasuda *et al.*, 2003; Gallo *et al.*, 2009). Susceptibillity to tricolsan, a broad-spectrum antimicrobial and anti-fungal agent, was also tested as it targets the FabI protein of FASII in bacteria (McMurry *et al.*, 1998; Heath *et al.*, 1999; Heath *et al.*, 2000). Triclosan has also demostrated antiparasitic activity in *Plasmodium* and *Trypanosoma* (Surolia and Surolia, 2001; Perozzo *et al.*, 2002; Paul *et al.*, 2004; Yu *et al.*, 2008a). However, the target of the drug in both parasites actually may not solely be the FabI enzyme *in vitro*, as was shown by Paul *et al*. (2004) in *T. brucei* and by Yu *et al*. (2008a) in *P. falciparum* and *P. berghei*.

Following 48 h incubation of WT 3D7 and 3D7$^{pCC1\text{-}\Delta PfaE3}$ clones 16 and 34 with BSO, paraquat or triclosan, the incorporation of [^{3}H]-hypoxanthine was used to determine IC_{50} values (section 2.9.12). No difference in IC_{50} values between WT 3D7 and the 3D7$^{pCC1\text{-}\Delta PfaE3}$ clones was observed for BSO or paraquat (Fig. 5-9A & B). Interestingly, a significant difference ($P < 0.001$) was seen in susceptibility to triclosan. 3D7$^{pCC1\text{-}\Delta PfaE3}$ clones were more resistant to triclosan (for both clones $IC_{50} = 14 \pm 1$ μM) than WT 3D7 ($IC_{50} = 8 \pm 1$ μM) (Fig. 5-9C). This is in contrast to the findings of Yu *et al*. (2008a), who showed that susceptibility of *P. falciparum* and *P. berghei FabI* deletion mutants to tricolsan did not differ from the WT parental strain. Due to time constrains, this experiment was only done once and thus needs to be repeated for further analysis.

A



B



C

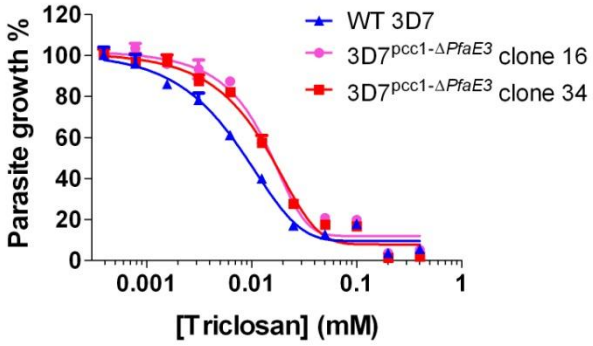


Figure 5-9. IC_{50} analyses of 3D7$^{pCC1\text{-}\Delta PfaE3}$ clones 16 and 34
Three technical replicates were set for each assay. The IC50 values are the mean ± standard error as determined using the log(inhibitor) vs response non-linear regression model in GraphPad Prism 5 (GraphPad Software Inc.). (A) BSO had no differing effect on the 3D7$^{pCC1\text{-}\Delta PfaE3}$ clones (IC50 = 82 ± 1 μM and 79 ± 1 μM clones 16 and 34, respectively) compared with WT 3D7 (IC50 = 82 ± 1 μM). (B) No effect was seen with paraquat (clone 16 IC50 = 12 ± 1 μM, clone 34 IC50 15 ± 1 μM, WT 3D7 IC50 20 ± 1 μM). (C) Triclosan IC50 for 3D7pCC1-ΔPfaE3 clone 16 (14 ± 1 μM) and clone 34 (14 ± 1 μM) was significantly higher than for WT 3D7 (8 ± 1 μM) ($P< 0.001$). The one-way analysis of variance (ANOVA) was used to calculate the statistical significance ($P< 0.05$ indicated signficance).

5.4 Summary

- The attempt to delete *PfE2* by double cross-over recombination using the pCC1 plasmid was not successful. Southern blot analyses showed integration of the plasmid, however, the *PfE2* loci was not targeted.

- *PfaE3* was successfully deleted in the 3D7 *P. falciparum* strain via double cross-over recombination using the pCC1 plasmid. Clonal lines homogenous for the mutation were obtained and named 3D7$^{pCC1\text{-}\Delta PfaE3}$.

- 3D7$^{pCC1\text{-}\Delta PfaE3}$ mutants showed a slightly impaired growth compared with WT 3D7 in the erythrocytic stages of the life cycle. The lack of *Pf*aE3 resulted in parasites that stayed highly synchronous compared with WT 3D7 parasties.

- Deletion of *Pf*aE3 had no effect on the susceptibility of the parasites to the oxidative stressors BSO or paraquat, however, the 3D7$^{pCC1\text{-}\Delta PfaE3}$ clones were more resistant to triclosan than the 3D7 wild type.

Chapter 6
Discussion

This study investigated the malaria parasite PDC, which is unusual in that it is only found in the apicoplast, the plastid-like organelle found in most apicomplexan parasites (Foth *et al.*, 2005; McMillan *et al.*, 2005). Unlike the mammalian PDC, which is located in the mitochondrion and produces acetyl-CoA for the citric acid cycle, the *Plasmodium* PDC produces acetyl-CoA for fatty acid biosynthesis. The PDC has been shown to be essential for liver stage parasites, as mutants lacking the E1 and E3 enzymes did not egress from hepatocytes and thus, could not elicit the intraerythrocytic stages of the malaria infection (Pei *et al.*, 2010). As the blood stages of the parasite life cycle cause the symptoms of malaria, the PDC could be a potential prophylactic drug target or vaccine candidate. Not much is known about the *Plasmodium* PDC proteins and if drug discovery is to progress it is necessary to be able to generate protein to investigate the structure and function of the complex and also ultimately study the effects of inhibitors.

6.1 *P. falciparum* protein expression and purification

Heterologous expression of *P. falciparum* proteins is plagued by difficulty, which is why obtaining high amounts of soluble proteins required for structural characterisation is often challenging or even impossible. Factors thought to cause poor expression include the high A/T content (80 %) of the *P. falciparum* genome and the high number of introns (Gardner *et al.*, 2002a). *Pf*E2 contains 10 introns and the A/T content of the mature protein is 73%. *Pf*aE3 does not contain introns but has an A/T content of 75 %. *P. falciparum* also prefers to use codons that are rare in *E. coli*, which can lead to depletion of the host tRNAs and halt expression (Saul and Battistutta, 1988; Kane, 1995; Kurland and Gallant, 1996). For example, the *P. falciparum* preferred codons for arginine (AGA/G), isoleucine (AUA) and glycine (GGA) are very rarely used in *E. coli* (Baca and Hol, 2000). A *P. falciparum* gene may also contain Shine-Dalgarno-like sequences resulting in incomplete protein translation (Cicek *et al.*, 2013). Both *Pf*E2 and *Pf*aE3 contained several rare codons and *Pf*aE3 also contained two Shine-Dalgarno-like sequences. Furthermore, *P. falciparum* proteins are usually larger than those from other species and contain long disordered stretches or loops (Pizzi and Frontali, 2001; Aravind *et al.*, 2003). This is the case for both proteins studied

here, with *Pf*E2 being about 8 kDa larger than other organisms and *Pf*aE3 being on average 12 kDa larger. Both proteins also contain longer unstructured loops than their counterparts from other organisms.

To obtain recombinant protein in this study, extensive optimisation of parameters was carried out including codon optimisation, choice of expression plasmid and *E. coli* host strain, expression temperature and medium, and the amount of IPTG required for induction were varied. Regardless, the final protein yields were low but enough to carry out meaningful analyses. Average yields from 1 L of bacterial culture were 1.8 mg for His-r*Pf*E2m, 0.25 mg for His-r*Pf*E2bc and 1.5 mg for His-r*Pf*aE3.

A few large-scale studies attempting heterologous expression of *P. falciparum* proteins in prokaryotic systems have been performed. These had the aim to generate recombinant protein for high throughput structural analyses potentially aiding drug discovery. The Structural Genomics of Pathogenic Protozoa group (SGPP, University of Washington) expressed 1000 *P. falciparum* open reading frames in *E. coli* (Mehlin *et al.*, 2006). They successfully expressed 33.7 % of the genes but only 6.3 % of these generated soluble proteins with yields between 0.9 – 406.6 mg/L of bacterial culture (Mehlin *et al.*, 2006). This just emphasises how variable and unpredictable the recombinant expression of proteins can be and that expressing *P. falciparum* proteins is particularly challenging. Upon closer analyses, a strong correlation of high isoelectric point (pI) was associated with poor expression of soluble recombinant protein; other factors such as size of the recombinant protein and greater structural flexibility were also contributing factors for low expression levels (Mehlin *et al.*, 2006). The pI of a protein was found to be especially important for expression of soluble recombinant proteins. The theoretical pI of His-r*Pf*E2m, 6.1, falls within in the optimum range for recombinant expression, while both His-r*Pf*E2bc and His-r*Pf*aE3 have a high predicted pI of 9.0 and 8.6, respectively, which may have contributed to the poor expression levels I found. The pI of a protein is influenced by the presence of basic and acidic amino acids and it was shown that proteins that expressed well had high quantities of aspartic acid, glutamic acid and histidine and low levels of tyrosine, lysine and arginine (Mehlin *et al.*, 2006). All three constructs studied here contained only 1.4 to 3.6% tyrosine and arginine residues, but they possessed 9.9 to 13.2% lysine residues. On the other hand, they contained few aspartic acid, glutamic acid and histidine residues (2.6-5.9 %), apart from His-r*Pf*E2m which contained 7.4 %

aspartic acid and 9.1 % glutamic acid. Therefore, proteins with higher levels of acidic residues have lower pI values, which could contribute to the better expression levels.

The estimated molecular masses of both His-r*Pf*E2m and His-r*Pf*aE3 (70 kDa and 64.5 kDa) are in the category of >60 kDa, which may have a negative effect on expression efficiency - the SGPP study found that only 20 % of proteins larger than 60 kDa expressed at all and of these only 3 % were found to be soluble (Mehlin *et al.*, 2006). This does not mean, however, that similar problems cannot be encoutnered with smaller proteins (Mehlin *et al.*, 2006) and indeed His-r*Pf*E2bc which has a predicted molecular mass of 34.5 kDa also showed a relatively low soluble expression. So the likelyhood is that other factors contribute to the poor expression levels I found for the proteins. For instance, *Pf*E2 contains highly disordered long linkers between the domains and *Pf*aE3 has several longer loops between structured elements, all of which could play a role in the poor expression of the proteins.

Another consortium (Structural Genomics Consortium SGC, Toronto, Canada), expressed 400 *P. falciparum* genes and also assessed expression of orthologues of these from other apicomplexan parasites (*P. yoelii, P. vivax, P. berghei, P. knowlesi, T. gondii* and *C. parvum*) (Vedadi *et al.*, 2007). From a total of 1008 genes, 30.2 % expressed as soluble proteins and this study concluded that protein size and pI did not correlate with expression levels, soluble expression and sucessful crystallisation of the proteins (Vedadi *et al.*, 2007).

Surprisingly, *P. falciparum* A/T content and codon bias was not shown to significantly affect expression (Mehlin *et al.*, 2006; Vedadi *et al.*, 2007). It could be reasoned that as codon optimisation would not change the amino acid sequence of the protein, it would not change the pI of the protein, thus it would not affect expression. However, there could be an effect of ribosome stalling due to limited availability of required amino acids, which could lead to errors and truncated proteins (Zhou *et al.*, 2004; Birkholtz *et al.*, 2008). Indeed, I found that truncated expression products were a problem in the expression profile of His-r*Pf*aE3, which contained the endogenous gene. However, truncated expression products were also seen with the codon optimised His-r*Pf*E2m. Several individual proteins have been expressed with better success by either codon optimising the whole gene or by co-expressing the gene in an *E. coli* strain with plasmids encoding for rare tRNAs (Baca and Hol, 2000; Zhou *et al.*, 2004; Dobano *et al.*, 2009; Evans *et al.*, 2012). A recombinant *Pf*E2m construct with the native sequence was not tested, thus it cannot be said whether

codon optimisation improved expression of His-r*Pf*E2m. The codon optimised construct, His-r*Pf*aE3-pET-15b, did not express soluble protein as well as the native construct, His-r*Pf*aE3. This, however, was more likely due to the codon optimised construct including 10 more amino acids at the N-terminus than the native construct. All or some of the additional residues in the codon optimised construct were most likely part of the transit peptide, thus affecting the solubility of the recombinant protein. For His-r*Pf*E2m expression with Rosetta™ 2(DE3)pLysS, which contains plasmids encoding rare tRNAs, was more successful but this was not the case for His-r*Pf*E2bc of His-r*Pf*aE3.

Another method related to codon optimisation is codon harmonisation, which has successfully improved *E. coli* expression of *P. falciparum* vaccine candidate proteins including merozoite surface protein 8 (*Pf*MSP8) (Alaro *et al.*, 2012), the transmission blocking vaccine target *Pf*s48/45 (Chowdhury *et al.*, 2009), liver-stage antigen 1 (LSA-1) (Hillier *et al.*, 2005) and FALVAC-1, a recombinant protein encoded by a synthetic gene consisting of epitopes from various parasite life-cycle stages (Zhou *et al.*, 2004). The frequency of a codon in an organism correlates with the amount of tRNA present, thus high frequency codons are translated faster than low frequency ones (Ikemura, 1981). Proteins begin their secondary structure formation already within the ribosomal tunnel during translation (Thanaraj and Argos, 1996). Therefore, the ribosome pausing at the low frequency codons is thought to regulate co-translational protein folding (Komar, 2009). Interestingly, high frequency codons are associated with the translation of α-helix structures, while low frequency codons are preferred by β-sheets and loops (Thanaraj and Argos, 1996). Hence, the placement of codons within an amino acid sequence is not random and codon harmonisation takes this into account. By substituting *P. falciparum* low frequency codons with *E. coli* low frequency codons and likewise for the high frequency codons, ribosomal processing can keep its native rhythm (Komar, 2009). Codon optimisation is more drastic and changes all the codons in the *P. falciparum* sequence, which are rare in *E. coli*, into codons that the host organism "prefers"; thus the ribosomal pauses required for correct folding may be lost. Therefore it may be worthwhile to codon harmonise the gene sequences encoding the recombinant proteins His-r*Pf*E2m, His-r*Pf*E2bc and His-r*Pf*aE3, which may have beneficial effects on the degree and levels of soluble expression in the prokaryotic system.

Other methods used to improve soluble expression of recombinant *P. falciparum* proteins were to add large, highly soluble proteins as fusion-protein tags at the N-terminus of the

protein of interest. Proteins such as glutathione-S-transferase (GST) (Das *et al.*, 2012), maltose binding protein (MBP) (Choveaux *et al.*, 2012) or thioredoxin (TRX) (Xiao *et al.*, 2007) were used with some success. The success-rate was between 5% and 16% when using MBP or GST-tagged constructs for the recombinant expression of *P. falciparum* proteins (Aguiar *et al.*, 2004). In this thesis, expression of the catalytic domain (CD) of *Pf*E2 was attempted using a fusion with GST. Expression of GST-r*Pf*E2c improved over His-r*Pf*E2bc, however, separation of the target protein, r*Pf*E2c, from the GST fusion partner following cleavage of the fusion protein was not achieved. r*Pf*E2c did not elute from the GST-FF purification column, suggesting r*Pf*E2c has an affinity to the glutathione coated beads or that it interacts or aggregated with the cleaved GST and was thus retained on the column. This is not unusual as Tsunoda *et al*. (2005) also reported difficulties following cleavage of a range of fusion tags (GST, MBP, TRX, His-tag) including aggregation, degradation, incorrect folding and precipitation of the target protein. Therefore, as with the other optimisation methods, expression and purification of *P. falciparum* proteins using fusion constructs is not a universally successful method. In fact, a survey of 57 potential drug targets from the WHO TDR Targets database (http://tdrtargets.org) showed constructs with His-tags to express better than the other large fusion tags (Birkholtz *et al.*, 2008). Thus, initial expression of a protein is worth testing first with His-tags rather than the more complex fusion proteins.

In some cases, purification and refolding of recombinant protein from inclusion bodies can be achieved to take advantage of the insoluble expression fractions. The target protein is often quite pure in the inclusion bodies, high protein concentrations can be obtained and protein in inclusion bodies is protected from proteolysis (Jungbauer and Kaar, 2007). The major downside is that the correct refolding of the protein following isolation from the inclusion bodies using denaturing purification methods may not be successful. Protein refolding is a complex process through several intermediate conformations, which can aggregate and precipitate (Jungbauer and Kaar, 2007). However, Foth *et al.* (2005) did purify *Pf*E2 CD using denaturing conditions followed by refolding of the protein and they obtained catalytically active protein, thus this method can be successful in some cases.

It is apparent that there is no simple solution to optimising *P. falciparum* protein expression in *E. coli* as there does not seem to be a unified set of parameters applicable to all soluble expressed proteins. Size, pI, amino acid composition and protein disorder play a role in some of the proteins. Codon optimisation or harmonisation is suitable for some

while co-expression with plasmids containing additional rare tRNAs may improve soluble expression of others. Expression of *P. falciparum* proteins with fusion partners can be successful in some cases. A recent development improving expression has been the co-expression the molecular chaperone, heat-shock protein 70, to aid correct folding of the *P. falciparum* recombinant protein (Stephens *et al.*, 2011). To improve the yield of the proteins studied here, codon harmonisation and co-expression with molecular chaperones is something that could be done in the future. Expression in other organisms such as yeast (LaCount *et al.*, 2009) and the baculovirus/insect cell system (Barfod *et al.*, 2006; Withers-Martinez *et al.*, 2012) could be tried in the future as these have been proven successful for a number of *P. falciparum* proteins. Also the cell-free wheat germ protein expression system has been successful in some cases (Iwanaga *et al.*, 2013).

Improving expression yield is not the only aspect that would improve the final amount of protein. In this study, protein aggregation and degradation during purification were also a major problem. There are several co-solvents that can be added to protein solutions to prevent formation of aggregates. Chaotropic agents (e.g. $MgCl_2$, $CaCl_2$) interact with peptide groups and hinder protein-protein interactions thus preventing aggregates, while kosmotropes (e.g. Na_2SO_4, $MgSO_4$, NaCl, KCl) help stabilise the native state of a protein (Bondos and Bicknell, 2003). Sugars, for example, glycerol can also stabilise proteins (Gekko, 1981). The reducing agents dithiothreitol (DTT) and β-mercaptoethanol prevent formation of unwanted disulphide bonds thus inhibiting aggregation (Bondos and Bicknell, 2003). Degradation on the other hand can be prevented by the addition of protease inhibitors during the protein extraction steps. EDTA is a metal chelator, which sequesters metals from metalloproteases, thus inhibiting them. Ideally during purification proteases will be separated form the target protein, thus eliminating proteolysis. Eliminating degradation due to other factors such as temperature and pH requires trial and error optimisation. A lot of aggregation and degradation was seen during purification of both His-r*Pf*E2m and His-r*Pf*aE3. This was partially improved by the addition of EDTA and increasing the concentration of NaCl. His-r*Pf*E2bc was very unstable and precipitated at concentrations higher than 0.7-0.8 mg/ml even with the addition of EDTA, glycerol and DTT. All the protein constructs studied here were more stable when stored at 4°C than at –20°C. Further work could be done to potentially improve the stability of these proteins by screening several other co-solvents and additives.

6.2 *P. falciparum* dihydrolipoamide acetyltransferase

6.2.1 Homology models of *P. falciparum* E2 domains show structural similarity to E2 from other organisms

To gain some insight into the structure of *Pf*E2, homology models of the *Pf*E2 domains were made using the I-TASSER server for protein structure and function predictions (http://zhanglab.ccmb.med.umich.edu/I-TASSER/, Zhang, 2008; Roy *et al.*, 2010). I-TASSER was chosen for the modelling as the server ranked first in the 10th critical assessment of techniques for protein structure prediction (CASP10) (http://predictioncenter.org/casp10/groups_analysis.cgi).

The I-TASSER homology models of *Pf*E2 domains were overall quite similar to structures from other organisms as evidenced by the low rms values (0.9-2.5Å). The *P. falciparum* first lipoyl domain (LD1) and the sub-unit binding domain (SBD) are more similar to those of the human E3BP, while the second lipoyl domain (LD2) and the catalytic domain (CD) were more similar to domains of bacterial counterparts. The residues on the *P. falciparum* SBD differ from those in other organisms, thus mutation studies would be an interesting future experiment to decipher which residues are involved in E1 and E3 binding. *P. falciparum* E2 is larger than several of its counterparts from other organisms. This is mainly attributable to longer linkers between the domains, especially the linker between LD2 and SBD, which is 37-56 amino acids longer than in E2 proteins from other organisms.

A trimer homology model of the *Pf*E2 CD was also generated based on the *A. vinelandii* and *G. strearothermophilus* crystal structures (Mattevi *et al.*, 1993a, Izard *et al*., 1999). Residues that could potentially be involved in intra-trimer interactions were mapped (Fig. 4-6). These residues are not conserved and most likely form different types of bonds in proteins from different organisms. The *Pf*E2 residues in these regions include several polar, hydrophobic and charged amino acids that could form hydrogen bonds and electrostatic interactions, required to form the trimer structure.

6.2.2 *P. falciparum* E2 does not readily form a large core

As in *A. vinelandii* and *G. strearothermophilus* (Mattevi *et al.*, 1992a; Izard *et al.*, 1999), *P. falciparum* E2 C-terminal residues form the 3_{10}-helix "ball", which should interact with a "socket" formed in the neighbouring trimer (Fig. 4-7). Based on this observation, the protein should be well adapted to form the high molecular mass core as is found for E2 from other organisms. Interestingly, however, His-r*Pf*E2m and His-r*Pf*E2bc do not seem to form the large E2 core readily. Fractionation of recombinant proteins by gel filtration resulted in the elution of several high molecular mass proteins, however, these were shown by AUC to be aggregates. An SV experiment on a concentration series of His-r*Pf*E2m showed a main species with $s^0_{20,w}$ = 6.6 ± 0.1 S (33-40% of sample), which is likely to correspond to a trimer of the protein. In addition, a minor species with $s^0_{20,w}$ = ~10.4 S (4-8 % of sample) was present at lower concentrations. This species may represent a hexamer of His-r*Pf*E2m. However, the expected 24mer or 60mer that E2 of other organisms readily form (Willms *et al.*, 1967; Danson *et al.*, 1979; Bosma *et al.*, 1984; Schulze *et al.*, 1991b; Stoops *et al.*, 1997; Lessard *et al.*, 1998; Vijayakrishnan *et al.*, 2010) was not detected. The only larger forms of His-r*Pf*E2m that were detected were protein aggregates - a total of 31-67 % of the sample. His-r*Pf*E2bc, on the other hand, contained only one species with $s^0_{20,w}$ = 4.5 ± 0.3 S and no aggregates.

These sedimentation coefficients suggest the main species is a trimer, which was confirmed by SE determination of the molecular mass of the main species. This was on average 217 kDa for His-r*Pf*E2m (monomer size 70 kDa) and 105 kDa for His-r*Pf*E2bc (monomer size 34.6 kDa). Mature-length *E. coli* E2 core, which forms a 24-mer, has a sedimentation coefficient of 23.1 S, while the value for a truncated core with only the CD is 17.1 S (Willms *et al.*, 1967; Schulze *et al.*, 1991b). *A. vinelandii* mature-length and truncated CD cores, also 24-mers, have sedimentation coefficients of 20.8 S and 17 S, respectively (Bosma *et al.*, 1984; Hanemaaijer *et al.*, 1987). The truncated 60-mer *G. stearothermophilus* core is 27.4 S (Packman *et al.*, 1984). The 60-mer potato mitochondrial PDC E2 had a sedimentation coefficient of 28 – 33 S depending on the concentration of the protein (Millar *et al.*, 1999). 60-mer mature-length human E2/E3BP core, was 29.3 S, while a truncated core consisting of the CD was 27.5 S (Vijayakrishnan *et al.*, 2010). No peaks corresponding to sedimentation coefficients as large as these were observed for His-r*Pf*E2m or His-r*Pf*E2bc.

The SE studies, however, indicated that there could be a large core complex present in the His-r*Pf*E2m sample. At the two highest concentrations used for this experiment, a 1.33-1.53 MDa species made up 11-16% of the sample. This larger species could represent a 24-mer, which theoretically should have a size of 1.67 MDa (based on the amino acid sequence). Interestingly, with decreasing concentrations of His-r*Pf*E2m larger proteins species were observed with sizes of 2 MDa and 3.7 MDa, which potentially correspond to a 30-mer and a 60-mer, respectively. These large molecular mass species, however, were not detected in SV studies. One possibility for this could be the high rotor speed (49,000 rpm) used in these experiments, which would sediment very large molecular mass species too fast for adequate analysis. Although, large aggregates with sedimentation coefficients above 40 S were seen in SV analyses at this speed as well.

The trend of increasing molecular mass with decreasing concentration seen in the SE experiments indicates thermodynamic non-ideality of the protein. The elongated shape of His-r*Pf*E2m affects the excluded volume of the protein, which in turn affects how the protein packs toward the base of the cell during an SE experiment. This further complicates the determination of the true molecular mass of the His-r*Pf*E2m multimers. Thus, the stoichiometry of the potential larger core complex is inconclusive and will require further studies using alternative approaches.

If *Pf*E2 does in fact form a large core complex, it can be implied that the interaction, at least *in vitro*, is not very strong. The E2 proteins of other organisms form the large core structure very readily regardless whether it is investigated using a recombinantly expressed protein or a purified native protein (Willms *et al.*, 1967; Danson *et al.*, 1979; Bosma *et al.*, 1984; Schulze *et al.*, 1991b; Stoops *et al.*, 1997; Lessard *et al.*, 1998; Vijayakrishnan *et al.*, 2010). The interactions required to form the large core between the E2 trimers involve two hydrogen bonds, thus it is a much weaker interaction than the intra-trimer interactions. This is thought to facilitate "breathing" of the core for optimum protein dynamics (Zhou *et al.*, 2001a). Despite this apparently "weak" interaction, to dissociate the 60-mer E2 core of *G. stearothermophilus,* 2 M guanidine hydrochloride was necessary. After reversing this dissociation, the 60 mer reassembled without any precipitation of the proteins and the re-folded protein retained its ability to bind E3 and was catalytically active (Lessard *et al.*, 1998). The 24-mer core of *E. coli* was also disassembled and it regained the correct core stoichiometry after refolding (Willms *et al.*, 1967). A factor that could weaken the interaction between the *P. falciparum* trimers is that the "socket" into which the C-terminal

3_{10}-helix "ball" is inserted, is predicted to be formed by one less residue than in other organisms as the helix H3 is shorter than that of E2 proteins from other organisms (Fig. 4.8).

Low protein concentration was another factor considered, which could affect core formation. The highest concentrations at which His-r*Pf*E2m and His-r*Pf*E2bc were assayed were 3.5 and 0.5 mg/ml, respectively. However, large complexes in other organisms have been observed by SV and electron microscopy (EM) experiments even at concentrations as low as 0.1 – 0.8 mg/ml (Willms *et al.*, 1967; Bosma *et al.*, 1984). Thus, the His-r*Pf*E2m and His-r*Pf*E2bc concentrations were in a range in which larger cores can form. Nevertheless, further studies into the effect of protein concentration on core formation could be done in the future. As obtaining a higher concentration of His-r*Pf*E2m is not possible, a "crowding experiment" may provide further insights into the structural peculiarities of r*Pf*E2m. This would be achieved by expressing the protein with an iLOV-tag (Light, Oxygen and Voltage), a fluorescent tag that can be spectrophotometrically tracked (Gawthorne *et al.* 2012). This recombinant protein would be mixed with a high concentration of another protein, for example, bovine serum albumin, to increase the general protein concentration of the solution. It is possible that the high protein concentration environment could encourage the interaction of iLOV-r*Pf*E2m trimers with each other, hence resulting in the formation of the large protein core.

The *A. vinelandii* PDC behaves slightly differently to the other complexes. The E2 forms a 20.8 S core on its own, however, upon binding of E1 and E3 the core dissociates into 18 S species, whereas the PDC of *E. coli* has a sedimentation coefficient of 60 S (Willms *et al.*, 1967; Bosma *et al.*, 1984). Determining the correct stoichiometry of the *A. vinelandii* complex was not as straight forward, as it was thought for years to be a 32-mer formed of eight tetramers (Bosma *et al.*, 1984). Experimental errors in determining the molecular mass of the monomer E2 were cited as the reasons for the incorrect stochiometry, which was instead found to be a 24-mer by Hanemaaijer *et al*. in 1987. The smaller building-block of the core was concluded to be a trimer once the crystal structure was solved (Mattevi *et al.*, 1992a). Thus, upon binding of E1 and E3 the *A. vinelandii* E2 24-mer core dissociates into trimers. These trimer PDCs are functional with about 2-fold higher specific activity compared with the bovine E2/E3BP 60-mer core (Schulze *et al.*, 1993; Yang *et al.*, 1997).

A small proportion of some *E. coli* PDC preparations have also shown an enzymatically active 19.8 S complex (Danson *et al.*, 1979). Thus, if *Pf*E2 was a trimer and did not form the expected larger core structure as other E2 proteins do, the protein still should have the ability to interact with E1 and E3 and form a functional PDC. Although the *A. vinelandii* E2 on its own forms a 24-mer core, the interactions were shown to be weaker than in other complexes as some dissociation into smaller units occurred even without the addition of E1 or E3 (Bosma *et al.*, 1984). Interestingly, the stoichiometry of the plant plastid E2 core is still unresolved as some dissociation occurs with this complex as well, resulting in broad variation in the measured molecular mass of the whole PDC (Camp and Randall, 1985; Olinares *et al.*, 2010). It has been suggested, however, that based on the ~ 5 MDa of the main PDC species, which is similar to that of the mitochondrial PDC protein, that the plant plastid E2 is a 60-mer (Mooney *et al.*, 2002; Olinares *et al.*, 2010).

Whether, in contrast with the trend observed for *A. vinelandii*, the addition of E1 and/or E3 encourages *Pf*E2 core formation will need to be assessed in future studies. In bacterial complexes it is thought that one E1 or E3 binds one SBD, however, recent studies with the human PDC have shown one E1 heterotetramer binds two E2 SBDs and one E3 dimer binds two E3BP SBDs (Smolle *et al.*, 2006; Vijayakrishnan *et al.*, 2010). Thus, addition of E1 and E3 could crosslink trimers to form a larger complex. Under some conditions the *A. vinelandii* PDC did form a large 52-60 S complex, which is comparable to the sedimentation coefficient of the *E. coli* PDC (Bosma *et al.*, 1984). This large complex was only seen in fresh protein preparations after the addition of PEG 6000 or 10 mM $MgCl_2$ or with increasing temperature (Bosma *et al.*, 1984). Testing whether the addition of these reagents affects the structure of His-r*Pf*E2m and His-r*Pf*E2bc would be a possibility, while increasing the temperature is not an option as both proteins are unstable at room temperature.

6.2.3 The linker between the *P. falciparum* E2 CD and SBD is extended

For the first time insights into the solution structure of an E2 trimer consisting of the CD and SBD has been obtained. Crystal structures have included only the CD and a portion of the linker, which in the bacterial structures extends over the neighbouring monomer in a hook-like conformation forming interactions thought to be involved in trimer formation (Mattevi *et al.*, 1992a; Izard *et al.*, 1999). In the human EM based CD structure, the linker

does not extend across in this manner (Yu *et al.*, 2008b). In the homology model of His-r*Pf*E2bc, which is based on the bacterial structures, the extension of the linker across the neighbouring monomer places the SBD in a position in which interaction with E1 and E3 would be sterically hindered (Fig. 4-9 and 4-10).

The EOM solution structure of the His-r*Pf*E2bc trimer suggests the presence of a flexible elongated linker that does not reach over the neighbouring monomer (Fig. 4-31). This is supported by E2 structures obtained from EM studies, where the linker also does not extend across the neighbouring monomer but reaches out from the first α-helix of the CD (Zhou *et al.*, 2001b; Gu *et al.*, 2003; Yu *et al.*, 2008b). From EM studies with *G. stearothermophilus* E2 in complex with E1 or E3, the length between the CD and the SBD was determined to be 75 - 95 Å (Milne *et al.*, 2002; Milne *et al.*, 2006). More constrained linker lengths of 50 – 80 Å were obtained from EM structures of mature-length human E2/E3BP and E1 bound yeast and bovine E2/E3BP (Zhou *et al.*, 2001b; Gu *et al.*, 2003; Yu *et al.*, 2008b). The *G. stearothermophilus* linker is 35 amino acids while the eukaryotic linkers are about 40 residues, which if allowing a 3.5 Å length for each residue, could stretch up to ~ 122 Å and ~140 Å (Zhou *et al.*, 2001b; Gu *et al.*, 2003). That the observed linker length is shorter than this suggests the linker has structured elements, for example, the human, bovine and yeast EM structures all have a linker with a bend in it (Zhou *et al.*, 2001b; Gu *et al.*, 2003; Yu *et al.*, 2008b). The *P. falciparum* linker between the CD and SBD is 31 amino acids long, thus it could stretch up to ~108 Å. Nine of the representative EOM models include linkers with lengths of 30 – 45 Å, which compare well to the lengths reported for the human, yeast and bovine linkers. Only one of the models had a linker of 60 Å. This suggests that the *P. falciparum* linkers have some structure to them as well and, in fact, a bend such as the one mentioned above can be seen in the EOM models (Fig. 4-31). Also, the estimated $s^{0}_{20,w}$ of 6.9 S for the compact homology model of His-r*Pf*E2bc does not correlate with the experimental values. The average estimated sedimentation coefficients for the elongated EOM models was 4.6 ± 0.1 S, which does correlate very well with the experimental value of 4.5 ± 0.3 S for His-r*Pf*E2bc, thus further validating this conformation.

The linkers were not extended in EM structures of truncated cores with only the CD and a portion of the linker, thus it was suggested that E1 binding was required to lengthen the linker (Zhou *et al.*, 2001a; Zhou *et al.*, 2001b; Yu *et al.*, 2008b). However, the results

shown here suggest that merely the presence of the SBD promotes extension of the linker. The longer linker lengths (50 - 95 Å) were obtained upon binding of E1, thus movement of the linkers seems to be required for PDC assembly and function (Gu *et al.*, 2003). It would be very interesting to obtain solution structures for His-r*Pf*E2bc with E1 and E3 to see the range of movement of which the linkers are capable of. The EOM structures grouped into two types of models, one with the linkers and SBD stretching out on the x/y-axis (Fig. 4-31A), the other with the linkers pointing upwards on the z-axis (Fig. 4-31B). If the *P. falciparum* PDC functions as a trimer, this movement would not seem to be a problem. However, if the protein forms a larger core complex, the most likely conformation would be with the linkers pointed upwards in the z-axis, as seen in the E2 EM structures. Any other conformation will most likely cause steric hindrance in E1 and E3 binding. For this reason it would be interesting to observe the effect of E1 and E3 binding on the conformation and placement of the linker and SBD with respect to the E2 trimer.

Some suggestions were made that E1 binds not only to the SBD but also to the CD (Schulze *et al.*, 1991a). Interaction of the CD core with E1 would not support the results obtained here or the EM structures, as this would not allow the extended conformation of the linker (Zhou *et al.*, 2001b; Milne *et al.*, 2002; Gu *et al.*, 2003). A mutation of *A. vinelandii* Arg416 in the CD (numbering based on *A. vinelandii* E2 sequence) to an aspartic acid weakened the binding of E1 and decreased the activity of the reconstituted PDC (Schulze *et al.*, 1991a). E1 interaction with the CD, however, was not seen in proteolysis studies of the CD (Packman *et al.*, 1984; Schulze *et al.*, 1991b) or in a more recent mutagenesis study where Arg416 was replaced by a glutamic acid (Chandrasekhar *et al.*, 2013). The role of the *A. vinelandii* CD-SBD linker was also probed by constructing a recombinant protein without the linker (Schulze *et al.*, 1991a). The deletion did not affect core formation (Schulze *et al.*, 1991a), thus the linker is not required to extend across the neighbouring monomer to form the core. E1 binding was lost, while E3 binding was not suggesting different binding sites for the two proteins. This does not mean that E1 necessarily binds to the CD. The construct was made so that the SBD was still present which, without the linker, would be right on top of the E2 trimer, thus it could be that the residues required for E1 binding were not accessible. Complete loss of E2 activity was also seen in the linker deletion mutant, which is most likely due to the deletion including a part of the first α-helix of the CD, which may be required for assembly of the E2 catalytic domain (Schulze *et al.*, 1991a). In a study where only the *A. vinelandii* linker was deleted but the first α-helix of the CD was left untouched, E2 catalytic activity was not lost,

neither was core formation (Schulze *et al.*, 1993). The lack of the linker did destabilise the trimer causing monomerisation upon E1 binding (Schulze *et al.*, 1993), however, again this could be due to the incorrect orientation of the SBD causing steric clashes resulting in monomerisation. If there is some interaction of E1 with the E2 CD, this could be a transient interaction required during PDC activity for correct substrate channelling. In conclusion, the linker does not need to extend over the neighbouring monomer for correct complex formation and activity. Thus, the extended linker model seen in this study and supported by the EM structures (Zhou *et al.*, 2001b; Gu *et al.*, 2003) is possible.

6.2.4 Catalytic activity of His-r*Pf*E2m and His-r*Pf*E2bc is higher than that of human E2

The specific activity of His-r*Pf*E2bc (80 ± 4 ΔAbs_{240mn}/min/mg) was 2.7 times higher than that of His-r*Pf*E2m (30 ± 0.9 ΔAbs_{240mn}/min/mg) (Table 3-3). This difference may well be attributable to the fact that the enzyme preparation (even after purification) of His-r*Pf*E2m contains inactive protein aggregates and degradation products, thus resulting in a decreased number of functional *Pf*E2 units. A similar difference was seen in *A. vinelandii*, where the truncated E2 CD core activity (117.9 ΔAbs_{240mn}/min/mg) was 1.5 times more active than the mature length protein (79.4 ΔAbs_{240mn}/min/mg) (Hanemaaijer *et al.*, 1989; Schulze *et al.*, 1991b). Compared with activity values reported by Foth *et al.* (2005) for the *Pf*E2 CD (153.8 ± 16.2 ΔAbs_{240mn}/min/mg), His-r*Pf*E2bc appears to be 1.9 times less active. Hanemaaijer *et al*. (1989b) reported a 1.6 times increase in *A. vinelandii* mature-length E2 activity following denaturing and refolding of the purified protein, suggesting some incorrect folding of the protein during expression or purification. However, purification methods have advanced since the 1980s as now the use of affinity tags has simplified the purification of soluble protein without the use of several precipitation steps, which could affect correct protein folding. Foth *et al*. (2005) purified the recombinant *Pf*E2 CD under denaturing conditions followed by refolding the protein. This may have removed interfering unfolded and aggregated protein species. However, the differences in specific activities observed in the two studies (this study and Foth et al., 2005) are not very large and may not be significant. However, denaturing purification is something that could be tried as several studies have denatured and refolded E2 successfully without losing activity or core formation ability (Willms *et al.*, 1967; Lessard *et al.*, 1998; Foth *et al.*, 2005).

His-r*Pf*E2bc activity was 4 times higher than that of the human E2 only core, 6.5 times higher than that of the bovine E2/E3BP core and 11 times higher than that of the *G. stearothermophilus* E2 core (Yang *et al.*, 1997; Lessard *et al.*, 1998, Table 3-3). The maize mitochondrial PDC E2 was the most active of all the organisms mentioned here, as it was 6.3 and 2.3 times more active than His-r*Pf*E2m and His-r*Pf*E2bc, respectively (Thelen *et al.*, 1999). The activity of the plant plastid E2 was measured with an assay different from the one used in this study, thus straightforward comparison with His-r*Pf*E2m and His-r*Pf*E2bc activities is not possible. However, the plant plastid E2 activity values measured by following ^{14}C-acetyl-CoA incorporation into acetyl-dihydrolipoamide, were similar to those for the human E2 and bovine E2/E3BP core (Yang *et al.*, 1997; Mooney *et al.*, 1999). The activity of His-r*Pf*E2bc was similar to that of *A. vinelandii* (Hanemaaijer *et al.*, 1989). This is interesting as *A. vinelandii* forms a functional trimer PDC. Apart from the maize mitochondrial PDC E2, the activities of both His-r*Pf*E2bc and *A. vinelandii* E2 are higher than those for organisms that function as 60- or 24-mer cores. Also the activity of the *E. coli* E2 is higher than that of the other large core forming organisms, and interestingly *E. coli* E2 can dissociate into smaller functional trimer PDC units (Danson *et al.*, 1979). Thus, it could be that if an E2 had to function as a trimer, the activity would need to be higher to assure adequate function of the complex. This would give further weight to the hypothesis that *Pf*E2 functions as a trimer.

Of the active site residues, His612 (*P. falciparum* full-length protein residue numbering) is the only strictly conserved amino acid. Ser561 is a serine in all other organisms apart from *G. stearothermophilus*, in which it is threonine, a residue with very similar properties to serine. His612 and Ser561 are the residues directly involved in the transfer of atoms in the catalytic reaction (de Kok *et al.*, 1998). In *A. vinelandii* Asn614 and Asp508 (*A. vinelandii* residue numbering) form a hydrogen bond in the inactive enzyme, thus preventing CoA entering and binding the active site (Fig. 1-11). These residues are not conserved in other organisms. In *P. falciparum* the corresponding residues are Tyr616 and Gly511. In other organisms these are an aspartic acid and an alanine or serine, respectively. Whether Tyr616 and Gly511 are important residues in the *P. falciparum* active site is something that could be further studied by mutational studies. It would be interesting to find out whether Tyr616 formed a bond with Gly511 or if it in fact interacted with either lysine residue flanking Gly511, which would be much more receptive to hydrogen bonding. In any case, activity probably is not affected by which residues form the inactive state bond as the variation in

these two residues in different organisms does not correlate with their differing activity values.

6.3 *P. falciparum* dihydrolipoamide dehydrogenase

The homology model of *Pf*aE3 shows the possible relative positioning of the FAD, NAD, central and interface domains (Fig. 4-32). The overall structure is unsurprisingly very similar to E3 proteins from other organisms with rmsd values ranging between 1.1.Å for the human E3 and 2.2 Å for the *A. vinelandii* E3. *Pf*aE3 is on average 12 kDa larger than its counterparts from other organisms. This extra size does not affect the structured elements of the protein as the majority is accounted for by six longer loops (Fig. 4-33).

SV analysis of His-r*Pf*aE3 shows the presence of a main peak with $s^{0}_{20,w}$ = 6.1 ± 0.1 S, which accounted for 68-94 % of the protein in the samples studied. SE analysis confirmed this to be on average a 120 kDa dimer (His-r*Pf*aE3 monomer size is 64 kDa). For comparison, the *A. vinelandii* E3 had a sedimentation coefficient of 5.9 S, while the values for human E3 range between 5.6-5.9 S (Hanemaaijer *et al.*, 1987; Hiromasa *et al.*, 2004; Smolle *et al.*, 2006; Brautigam *et al.*, 2009; Vijayakrishnan *et al.*, 2010). A ~4.1 S degradation product, with a molecular mass corresponding to a truncated monomer (60-62 kDa), accounted for 4-18 % of the samples. The degradation was thought to have occurred from the C-terminus, thus preventing dimerisation. Confirmation of this could be done by creating various *Pf*aE3 constructs truncated from the C-terminus to see how many residues are required for dimerisation. In three out of six samples 2-9 % of the sample was accounted for by a ~8.8 S species. Following SE analysis to obtain molecular mass, these were concluded to be 233-238 kDa tetramers.

The solution structure of His-r*Pf*aE3 shows the structure to be more elongated than the homology model based on the crystal structure of E3 from other organisms (Fig. 4.41). As the protein is highly conserved the crystal structures are very similar and have D_{max} values of approximately 110 Å (Chandrasekhar *et al.*, 2013). The D_{max} of His-r*Pf*aE3 determined from the distance distribution function, *p(r)*, was 133 Å, while the D_{max} of the final *ab initio* model was 146-149 Å. An elongated, more flattened E3 with a D_{max} of 130 Å has also been observed for the human E3 (Smolle *et al.*, 2006). Also, the sedimentation

coefficients suggest a less compact and elongated protein as the $s^{0}_{20,w}$ computed for the His-r*Pf*aE3 homology model was 7.35 S, which is much higher than the experimentally obtained values for His-r*Pf*E3 (6.1 ± 0.1 S). Surprisingly, however, the computed sedimentation coefficient for the DAMMIF model, 6.8 S, was still higher than the experimental value even though the shapes of the models are more elongated. The long loops and flexible regions between the domains of *Pf*aE3 could affect protein sedimentation and thus lower the experimental sedimentation coefficient. Flexibility of a protein can result in an incorrect DAMMIF model leading to the higher computed sedimentation coefficient, thus EOM and DMD were used to model the flexible His-tag at the N-terminus of His-r*Pf*aE3. This flexible region did seem to account for some of the empty space in the DAMMIF model (Fig. 4-44). However, the short 14-amino acid N-terminus region was not substantial enough to cause a decrease in the computed sedimentation coefficient (7.0 – 7.3 S). Thus, the extended and less compact structure of His-r*Pf*aE3 could to be due to loops between the domains, which would need to be modelled in the future.

The active site of *Pf*aE3 consists of the same conserved residues as in all other organisms; the two cysteines (Cys165 and Cys170), His646 and Glu 651 (Fig. 4-34). In the reverse reaction, NADH and lipoamide K_m and V_{max} values for His-r*Pf*aE3 (Table 3-5) were comparable with those reported previously (McMillan, 2006). The values are similar to those for the *P. falciparum* mitochondrial E3 (*Pf*mE3) and other organisms (Kim *et al.*, 1991; Schöneck *et al.*, 1997; McMillan *et al.*, 2005). The k_{cat} of the reverse reaction for the apicoplast was determined to be 10 times lower than that for *Pf*mE3 (McMillan, 2006). The plant chloroplast E3 was also shown to be 5 times less active than the mitochondrial counterpart (Conner *et al.*, 1996). In this study, the K_m and V_{max} values for the forward reaction substrates NAD^+ and DHLA were also determined. The affinity of His-r*Pf*aE3 for DHLA was 8 times lower than that for the mitochondrial E3, while the affinity for NAD^+ was about 5 times higher. A 6-fold higher affinity for NAD^+ has also been reported for the *Pisum sativum* chloroplast E3 compared with its mitochondrial E3 (Conner *et al.*, 1996). It was also shown that, unlike the mitochondrial E3, the chloroplast protein did not show any activity with $NADP^+$ (Conner *et al.*, 1996). It was concluded that low $NADP^+$ activity was required because, unlike in the mitochondrion, the levels of NAD(H) are lower than those of NADP(H) in the chloroplast stroma, thus requiring higher affinity of the chloroplast E3 to NAD^+ (Conner *et al.*, 1996). Whether His-r*Pf*aE3 is active using $NADP^+$ is something

that needs to be tested to further understand the environment in which the apicoplast PDC functions.

6.4 E2 and E3 gene deletion from *P. falciparum*

Gene manipulation in *P. falciparum* is very challenging due to the A/T rich genome resulting in difficulty with PCR generation of inserts for the knockout plasmids. Also transfection of *P. falciparum* is of very low efficiency, in the order of 1 in 10^6 (O'Donnell *et al.*, 2002), which makes obtaining transfectants time consuming. The low efficiency is likely due to the fact that transfection can only be done with ring stage parasites, thus the DNA has to cross four membranes (erythrocyte membrane, parasitophorous vacuole, parasite cell membrane, nucleus) before being able to integrate into the genome. In this study positive/negative drug selection using the pCC1 plasmid was used to obtain deletion of the target gene with double crossover recombination (Maier *et al.*, 2006).

Deletion of *PfE2* from *P. falciparum* was unsuccessful as the correct gene locus was not targeted. The *Pf*E2 gene contains 10 A/T rich introns, thus designing the gene deletion cassette was very difficult. The pCC1-Δ*Pf*E2 construct was intended to delete 1941 bp from the centre of the 3192 bp *Pf*E2 gene. It could be that this deletion was too short for specific targeting of the *Pf*E2 locus. Also the flanks used for targeting the gene could have been too short for specific targeting of *PfE2*. Flanks up to 1000 bp could be tried in future experiments. However, the failed attempt to delete *PfE2* could also be because the gene is required for *P. falciparum* survival and thus could not be targeted. Further attempts to delete *Pf*E2 will be made using a different approach with the recently published DiCre system, which allows inducible excision of genes from the *P. falciparum* genome (Collins *et al.*, 2013).

The deletion of the *PfaE3* was successful, thus the *Pf*aE3 protein is not essential for *P. falciparum* erythrocytic stages implying that PDC is not essential for these parasite stages. However, interest in potential other roles of *Pf*aE3 in the parasite warranted further investigation of the parasite line. $3D7^{pCC1\text{-}\Delta PfaE3}$ showed a growth phenotype, which was unexpected. The mutant parasites progressed through their erthrocytic cell cycle slower than the WT 3D7 parasites. Also, $3D7^{pCC1\text{-}\Delta PfaE3}$ remained highly synchronous for 3 cycles

(7 days), while wild type 3D7 parasites lost synchronicity after 2 cycles (4 days). The deletion of E1 or E3 in *P. yoelii* did not affect growth of blood stage parasites in mice (Pei *et al.*, 2010). It is possible that experimental differences caused the mild phenotype observed here while this was not obvious in the mouse model. This is reminiscent of the varying observations of the growth phenotype of a variety of gene deletions in *P. falciparum* and mouse malaria species. The deletion of *lipB* in *P. falciparum* resulted in a shorter cell cycle phenotype (the mutant parasites progressed about 8 hours faster through the intra-erythorcytic cycle than the wild type parasites), which was not seen in the *P. berghei lipB* deletion mutant (Günther *et al.*, 2007; Falkard *et al.*, 2013). Deletion of the FAS II enzymes FabI, FabB/F and FabZ did not show any effect on *P. falciparum* intra-erythrocytic growth *in vitro* or *P. yoelii* growth *in vivo* (Yu *et al.*, 2008a; Vaughan *et al.*, 2009). Thus, it could be that the growth phenotype seen in this study (and the *lipB* study) is due to an effect other than disruption of PDC and independent of FAS II. Acetylation is a major posttranslational modification, which is involved in and regulates a vast range of cellular processes including the cell cycle (Choudhary *et al.*, 2009). Thus, potentially a decrease in acetyl-CoA levels followed by reduced acetylation of proteins involved in the regulation of the cell cycle could cause the growth phenotype. In any case, the growth assay will need to be repeated and further studies on the 3D7$^{pCC1\text{-}\Delta PfaE3}$ mutant will be carried out to better understand the effect *Pf*aE3 deletion is having on the parasites.

Susceptibility of the 3D7$^{pCC1\text{-}\Delta PfaE3}$ mutant lines was tested against oxidative stressors and the FASII inhibitor triclosan. Susceptibility to the oxidative stressors, BSO and paraquat, did not differ from that of the 3D7 WT lines. BSO inhibits the glutathione synthesis pathway, a major redox balance mechanism in the cytoplasm (Patzewitz *et al*., 2012) and paraquat is a general ROS producer. Thus, reducing equivalents from the apicoplast, which is separated from the cytosol by four membranes, would most likely not have a significant effect on combating oxidative stress in the cytoplasm caused by these agents. Thus, *Pf*aE3 does not play a role in oxidative defence of the organelle. A significant increase ($P< 0.001$) in resistance to triclosan was determined in the 3D7$^{pCC1\text{-}\Delta PfaE3}$ mutants. It has been shown that FabI is not the only target for triclosan and that the drug may also have an effect on for instance parasite membranes (Paul *et al.*, 2004; Yu *et al.*, 2008a). Why the 3D7$^{pCC1\text{-}\Delta PfaE3}$ mutants would be more resistant to triclosan is unclear. Due to time constraints, these experiments were done only once, therefore before launching into further speculation, the experiments need to be repeated to verify whether the difference is truly significant or not.

6.5 Conclusions and future perspectives

Expression of soluble *P. falciparum* protein is challenging and purification of native or recombinant proteins can be taxing due to instability of the proteins or simply the availability of sufficient biological materials when working with native proteins. In this study several methods were employed to improve expression of recombinant proteins and their purification to allow subsequent structural analyses. *Plasmodium* gene sequences were *E. coli* codon optimised, cloned into various expression vectors, grown in a range of *E. coli* host strains at different temperatures, in different growth media for various time periods and at various IPTG concentrations. In addition the length of the expressed proteins was varied to improve their expression and purification. Improving purification of the recombinant proteins was assessed by supplementing buffers with several additives known to prevent degradation and aggregation such as EDTA, glycerol and DTT. Overall, this resulted in the successful expression, purification and concentration of the mature-length *P. falciparum* E2, His-r*Pf*E2m up to a yield of 1.8 mg//L of bacterial culture. The truncated E2, His-r*Pf*E2bc, containing the CD and SBD could be obtained only at 0.25 mg/L of bacterial culture due to instability of the protein. For mature-length *P. falciparum* E3, His-r*Pf*E3, maximum final yield was 1.5 mg/L of bacterial culture. Due to extensive optimisation of expression in the *E. coli* host performed here, the next steps to improve this would be to try other expression systems. These could include expression in yeast, insect cells or the wheat germ cell-free system (Barfod *et al.*, 2006; LaCount *et al.*, 2009; Iwanaga *et al.*, 2013). It would be interesting to obtain His-r*Pf*E2bc at higher concentrations so that crystallisation could be attempted. A more stable mature-length *P. falciparum* E2, which does not aggregate, would enable interesting SAXS studies on the mature length trimer to complement results obtained for His-r*Pf*E2bc. Also a more stable E3 construct would be useful in crystallisation studies.

The most interesting finding of this study was that *P. falciparum* E2 does not readily form a large 24-mer or 60-mer core as seen for E2 proteins of other organisms. Some indication of a larger complex was seen in SE experiments with the mature-length *Pf*E2, however, these larger species were not observed in SV experiments. Also, no indication of a larger complex was seen with the truncated His-r*Pf*E2bc. Both native and recombinant E2 from bacteria, yeast, bovine and humans form the large central core even at concentrations below 1 mg/ml and even after denaturing and refolding (Willms *et al.*, 1967; Danson *et al.*,

1979; Bosma *et al.*, 1984; Schulze *et al.*, 1991b; Stoops *et al.*, 1997; Lessard *et al.*, 1998; Vijayakrishnan *et al.*, 2010). Additional studies need to be done to further probe this. One option is to perform a "crowding experiment", where *Pf*E2 is placed into a high protein concentration environment to see whether this would increase the affinity of E2 trimers for each other and drive core formation. Another possibility would be to attempt purification of native *Pf*E2 from the parasites. This would be difficult as the PDC proteins are expressed only at low levels during the erythrocytic stages, thus large amounts of parasites are required to obtain sufficient starting material for purification. The Müller lab does have an antibody reactive with *Pf*E2, which will be used to immunoprecipitate the protein from parasite extracts. Using this protein it may be possible to determine the stoichiometry of *Pf*E2 using native PAGE or mass spectrometry. Also EM could be carried out to obtain images of the *Pf*E2 core. Another advantage when working with the native E2 protein is that the protein may have E1 and E3 bound to it, thus enabling the study of the full native PDC complex in structural and kinetic studies. Another experiment that will be performed in the future is to assess whether the addition of E3 affects E2 core formation. The effect of both E1 and E3 will require the expression and purification of the two E1 subunits. This is likely to be complicated by the fact that E1 is formed by two subunits of E1α- and two subunits of E1β-domains, which probably need to be co-expressed to generate the fully active and properly folded E1-protein. But if this can be achieved, studies can be done to elucidate the whole *P. falciparum* PDC function.

The lack of large core formation permitted, for the first time, the acquisition of solution structural information for an E2 trimer was obtained. The *Pf*E2 SBD is separated from the CD by a 30 – 45 Å linker, which correlates well with EM studies that show a 50 – 80 Å gap between the CD and bound E1 in the yeast, human and bovine complexes (Zhou *et al.*, 2001b; Gu *et al.*, 2003; Yu *et al.*, 2008b). The linker in these organisms is approximately 10 residues longer, which could account for the shorter gap between the *P. falciparum* CD and SBD. Also the binding of E1 or E3 could further extend the linker (Gu *et al.*, 2003). This is in contrast to the bacterial crystal structures, where a major portion of the linker extends over the neighbouring monomer and is thought to be involved in intra-trimer interactions. However, mutation studies have shown that the linker does not need to extend over the monomer for core formation or activity (Schulze *et al.*, 1993). Also the sedimentation coefficient of His-r*Pf*E2bc supports the extended linker model and not the compact homology model based on the crystal structures of E2 from other organisms. In

the future it would be very interesting to carry out SAXS experiments with added E1 and E3 to see how the linker dynamics are affected.

The first major difference between *P. falciparum* and human E2 is that the human E2 core is formed of trimers comprising two E2 monomers and one E3BP or three E2 monomers (Vijayakrishnan *et al.*, 2010; Vijayakrishnan *et al.*, 2011). *P. falciparum* does not possess E3BP and therefore forms trimers with three E2 monomers. A further important difference may be the lack of multimer formation beyond a trimer implying that *Pf*E2 may be catalytically active as a trimer, which suggests that PDC in the parasite apicoplast has a very divergent structure to those of other PDCs. But even if a large multimeric core was formed, it still would seem that dissociation into the trimeric building blocks may occur readily – thus the structure of the PDC core may well be much less stable than that of, for example, the human enzyme. If this were indeed the case, it would lead to exposure of protein surface that may be exploited in drug discovery especially as in the human PDC equivalent areas would be buried inside the core.

It appears that the *Pf*E2 trimer is more active than its human counterpart, and interestingly has an activity similar to that of *A. vinelandii* E2, which was shown to function as a trimer PDC (Hanemaaijer *et al.*, 1989). Whether there is a correlation between higher E2 activity and full PDC activity would need to be studied. The *A. vinelandii* trimer PDC is twice as active as the bovine E2/E3BP 60-mer core (Schulze *et al.*, 1993; Yang *et al.*, 1997), thus it could be that a trimer *P. falciparum* PDC would also be more active than its mammalian 60-mer counterparts. The residues involved in catalytic activity, His612 and Ser561 (full-length *Pf*E2 numbering), are conserved between the parasite and human proteins. However, the two residues thought to be involved in a hydrogen bond to form the inactive state conformation, Tyr616 and Gly 511, are an aspartic acid and serine in human E2. A drug that would prevent the conformational change required for the formation of the active catalytic site in *P. falciparum* but not in the human enzyme could be a possibility. Residues in helix H1 of the E2 CD are also required for correct active site formation (Schulze *et al.*, 1991a). This is something that could be further studied with mutational studies, as these residues are not conserved between organisms and could differ enough to be exploited in drug development.

The final interesting difference is the binding of E1 and E3. Human E3BP binds E3 and the E2 SBD binds E1. In *P. falciparum* the E2 SBD binds both E1 and E3. In the *P.*

falciparum SBD homology model the amino acids in the positions of residues known to be involved in *G. stearothermophilus* E1 and E3 binding are very different (Mande *et al.*, 1996; Jung *et al.*, 2003; Jung and Perham, 2003; Frank *et al.*, 2005, Fig. 4-3, Table 4-1). This suggests *P. falciparum* may have a different mechanism of binding the peripheral enzymes. Furthermore, several residues involved in E3 binding to E3BP are very different to the *P. falciparum* residues (Brautigam *et al.*, 2006; Ciszak *et al.*, 2006, Table 4-2). Also the E2 amino acids involved in human E1 binding are thought to be similar to the *G. stearothermophilus* residues (Korotchkina and Patel, 2008). Thus, future structural and mutational studies of *Pf*E2 binding of E1 and E3 would be very worthwhile as a sufficient difference in binding the two proteins could also be exploited in inhibiting *P. falciparum* PDC activity.

As expected, *Pf*aE3 is a homodimer as in other organisms. Its hydrodynamic parameters and solution structure showed that the shape of the protein is more elongated and it is not as compact as homology models based on crystal structures of E3 from other organisms. The more elongated shape of *Pf*aE3 has also been observed for human E3 and is supported by SV values obtained for proteins of other organisms (Hanemaaijer *et al.*, 1987; Hiromasa *et al.*, 2004; Smolle *et al.*, 2006; Brautigam *et al.*, 2009; Vijayakrishnan *et al.*, 2010). *Pf*aE3 is approximately 12 kDa larger than human E3. However, the extra size is distributed in loops rather than the structured elements of the protein, thus the homology model of *Pf*aE3 actually had the highest similarity to the human protein (rmsd = 1.1 Å). Crystallisation of *Pf*aE3 could be attempted with the protein concentrations that can already be obtained. *Pf*aE3 is less active than either its mitochondrial counterpart or human E3 (Kim *et al.*, 1991; McMillan, 2006). *Pf*aE3 also has a higher affinity for NAD^+ than the *P. falciparum* mitochondrial E3 or human E3 (Kim *et al.*, 1991; McMillan, 2006). Whether this is something that could be exploited for drug development needs to be assessed in further studies. Although designing a drug that would be specific to *Pf*aE3 would most likely be difficult due to the high structural similarities not only to the homologous protein in the human host but also the structural and functional similarlty of E3 to other NAD(P)H dependent FAD disulphide oxidoreducatses (Petsko, 1991; Perham *et al.*, 1996; Chandrasekhar *et al.*, 2013).

Much more work will need to be done to characterise the $3D7^{pCC1\text{-}\Delta PfaE3}$ mutants and it would be interesting to try an alternative strategy to delete *Pf*E2 from the parasites. If *Pf*E2 is truly essential to the parasite, questions arise why *Pf*E2 is required but *Pf*E3 is not. One

possibility could be that another oxidoreductase or a related reducing enzyme was substituting the function of *Pf*E3. For example, a glutathione reductase and a glutathione peroxidase-like thioredoxin peroxidase have been shown to be targeted to the apicoplast (Kehr *et al*., 2010).

Manipulating the *P. falciparum* genome is difficult and very time consuming, thus the 3D7$^{pCC1\text{-}\Delta PfaE3}$ mutants were obtained only at the very end of laboratory work, and hence only a few selected assays were carried out once. The growth assay will need to be repeated to verify the interesting growth phenotype. Additionally a growth assay with shorter time points over two cycles would be useful to pin point the stage which causes the slower growth of the mutant parasites. Also the number of merozoites produced by the 3D7$^{pCC1\text{-}\Delta PfaE3}$ mutants compared with the wild type could be counted to see whether the lower multiplication rate is due to fewer merozoites being produced. The susceptibility test that implies a greater sensitivity to triclosan needs to be repeated. To further probe the role of the PDC in the blood stages, the parasites could be grown in media without fatty acids to see whether the mutant parasites will survive when fatty acids are not readily available from the medium. Targeted metabolomics will be carried out with the mutant parasites to study the effects that the deletion of *Pf*aE3 might have on parasite metabolism. It would be particularly interesting to assess the levels of acetyl-CoA, and to analyse whether compensatory reactions are induced through the removal of the PDC activity.

This study on *P. falciparum* E2 and E3 has uncovered fascinating aspects of the PDC, which could substantially differ from the human complex, thus making the PDC a candidate drug target. This work provdes a fantastic starting point and platform from which to continue further dissection of not only the interesting stoichiometry and assembly of the complex but the role of the PDC in *P. falciparum* and other *Plasmodium* parasites.

References

Abkarian, M., G. Massiera, L. Berry, M. Roques and C. Braun-Breton (2011). "A novel mechanism for egress of malarial parasites from red blood cells." *Blood* **117**(15): 4118-4124.

Aevarsson, A., J. L. Chuang, R. M. Wynn, S. Turley, D. T. Chuang and W. G. Hol (2000). "Crystal structure of human branched-chain alpha-ketoacid dehydrogenase and the molecular basis of multienzyme complex deficiency in maple syrup urine disease." *Structure* **8**(3): 277-291.

Aguiar, J. C., J. LaBaer, P. L. Blair, V. Y. Shamailova, M. Koundinya, J. A. Russell, F. Huang, W. Mar, R. M. Anthony, A. Witney, S. R. Caruana, L. Brizuela, J. B. Sacci, Jr., S. L. Hoffman and D. J. Carucci (2004). "High-throughput generation of *P. falciparum* functional molecules by recombinational cloning." *Genome Res* **14**(10B): 2076-2082.

Aikawa, M. (1966). "The fine structure of the erythrocytic stages of three avian malarial parasites, *Plasmodium fallax*, *P. lophurae*, and *P. cathemerium*." *Am J Trop Med Hyg* **15**(4): 449-471.

Alaro, J. R., E. Angov, A. M. Lopez, H. Zhou, C. A. Long and J. M. Burns, Jr. (2012). "Evaluation of the immunogenicity and vaccine potential of recombinant *Plasmodium falciparum* merozoite surface protein 8." *Infect Immun* **80**(7): 2473-2484.

Allary, M., J. Z. Lu, L. Zhu and S. T. Prigge (2007). "Scavenging of the cofactor lipoate is essential for the survival of the malaria parasite *Plasmodium falciparum*." *Mol Microbiol* **63**(5): 1331-1344.

Aravind, L., L. M. Iyer, T. E. Wellems and L. H. Miller (2003). "*Plasmodium* biology: genomic gleanings." *Cell* **115**(7): 771-785.

Arnot, D. E. and K. Gull (1998). "The *Plasmodium* cell-cycle: facts and questions." *Ann Trop Med Parasitol* **92**(4): 361-365.

Baca, A. M. and W. G. Hol (2000). "Overcoming codon bias: a method for high-level overexpression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*." *Int J Parasitol* **30**(2): 113-118.

Baer, K., C. Klotz, S. H. Kappe, T. Schnieder and U. Frevert (2007). "Release of hepatic *Plasmodium yoelii* merozoites into the pulmonary microvasculature." *PLoS Pathog* **3**(11): e171.

Baker, D. A. (2010). "Malaria gametocytogenesis." *Mol Biochem Parasitol* **172**(2): 57-65.

Bannister, L. H., J. M. Hopkins, R. E. Fowler, S. Krishna and G. H. Mitchell (2000). "A brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages." *Parasitol Today* **16**(10): 427-433.

Barfod, L., M. A. Nielsen, L. Turner, M. Dahlback, A. T. Jensen, L. Hviid, T. G. Theander and A. Salanti (2006). "Baculovirus-expressed constructs induce immunoglobulin G that recognizes VAR2CSA on *Plasmodium falciparum*-infected erythrocytes." *Infect Immun* **74**(7): 4357-4360.

Bartfai, R., W. A. Hoeijmakers, A. M. Salcedo-Amaya, A. H. Smits, E. Janssen-Megens, A. Kaan, M. Treeck, T. W. Gilberger, K. J. Francoijs and H. G. Stunnenberg (2010). "H2A.Z demarcates intergenic regions of the *Plasmodium falciparum* epigenome that are dynamically marked by H3K9ac and H3K4me3." *PLoS Pathog* **6**(12): e1001223.

Bauer, J. H. and E. G. Pickels (1937). "An improved air-driven type of ultracentrifuge for molecular sedimentation." *J Exp Med* **65**: 565-586.

Benen, J., W. van Berkel, C. Veeger and A. de Kok (1992). "Lipoamide dehydrogenase from *Azotobacter vinelandii*. The role of the C-terminus in catalysis and dimer stabilization." *Eur J Biochem* **207**(2): 499-505.

Berg, A., J. Vervoort and A. de Kok (1996). "Solution structure of the lipoyl domain of the 2-oxoglutarate dehydrogenase complex from *Azotobacter vinelandii*." *J Mol Biol* **261**(3): 432-442.

Berg, A., J. Vervoort and A. de Kok (1997). "Three-dimensional structure in solution of the N-terminal lipoyl domain of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*." *Eur J Biochem* **244**(2): 352-360.

Bernado, P., E. Mylonas, M. V. Petoukhov, M. Blackledge and D. I. Svergun (2007). "Structural characterization of flexible proteins using small-angle X-ray scattering." *J Am Chem Soc* **129**(17): 5656-5664.

Biewenga, G. P., G. R. Haenen and A. Bast (1997). "The pharmacology of the antioxidant lipoic acid." *Gen Pharmacol* **29**(3): 315-331.

Birkholtz, L. M., G. Blatch, T. L. Coetzer, H. C. Hoppe, E. Human, E. J. Morris, Z. Ngcete, L. Oldfield, R. Roth, A. Shonhai, L. Stephens and A. I. Louw (2008). "Heterologous expression of plasmodial proteins for structural studies and functional annotation." *Malar J* **7**: 197.

Blanchet, C. E., A. Zozulya, A. Kikhney, D. Franke, P. Konarev, W. Shang, R. Klaering, B. Robrahn, C. Hermes, F. Cipriani, D. Svergun and M. Roessle (2012). "Instrumental setup for high-throughput small- and wide-angle solution scattering at the X33 beamline of EMBL Hamburg." *J Appl Cryst* **45**: 489-495.

Block, H., B. Maertens, A. Spriestersbach, N. Brinker, J. Kubicek, R. Fabis, J. Labahn and F. Schafer (2009). "Immobilized-metal affinity chromatography (IMAC): a review." *Methods Enzymol* **463**: 439-473.

Bondos, S. E. and A. Bicknell (2003). "Detection and prevention of protein aggregation before, during, and after purification." *Anal Biochem* **316**(2): 223-231.

Borst, P., J. P. Overdulve, P. J. Weijers, F. Fase-Fowler and M. Van den Berg (1984). "DNA circles with cruciforms from *Isospora* (*Toxoplasma*) *gondii*." *Biochim Biophys Acta* **781**(1-2): 100-111.

Bosma, H. J., A. De Kok, B. W. Van Markwijk and C. Veeger (1984). "The size of the pyruvate dehydrogenase complex of *Azotobacter vinelandii*. Association phenomena." *Eur J Biochem* **140**(2): 273-280.

Botte, C. Y., Y. Yamaryo-Botte, J. Janouskovec, T. Rupasinghe, P. J. Keeling, P. Crellin, R. L. Coppel, E. Marechal, M. J. McConville and G. I. McFadden (2011). "Identification of plant-like galactolipids in *Chromera velia*, a photosynthetic relative of malaria parasites." *J Biol Chem* **286**(34): 29893-29903.

Boumis, G., G. Giardina, F. Angelucci, A. Bellelli, M. Brunori, D. Dimastrogiovanni, F. Saccoccia and A. E. Miele (2012). "Crystal structure of *Plasmodium falciparum* thioredoxin reductase, a validated drug target." *Biochem Biophys Res Commun* **425**(4): 806-811.

Bozdech, Z., M. Llinas, B. L. Pulliam, E. D. Wong, J. Zhu and J. L. DeRisi (2003). "The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*." *PLoS Biol* **1**(1): E5.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." *Anal Biochem* **72**: 248-254.

Brautigam, C. A., J. L. Chuang, D. R. Tomchick, M. Machius and D. T. Chuang (2005). "Crystal structure of human dihydrolipoamide dehydrogenase: NAD+/NADH binding and the structural basis of disease-causing mutations." *J Mol Biol* **350**(3): 543-552.

Brautigam, C. A., R. M. Wynn, J. L. Chuang, M. Machius, D. R. Tomchick and D. T. Chuang (2006). "Structural insight into interactions between dihydrolipoamide dehydrogenase (E3) and E3 binding protein of human pyruvate dehydrogenase complex." *Structure* **14**(3): 611-621.

Brautigam, C. A., R. M. Wynn, J. L. Chuang and D. T. Chuang (2009). "Subunit and catalytic component stoichiometries of an in vitro reconstituted human pyruvate dehydrogenase complex." *J Biol Chem* **284**(19): 13086-13098.

Brookes, E., B. Demeler and M. Rocco (2010a). "Developments in the US-SOMO bead modeling suite: new features in the direct residue-to-bead method, improved grid routines, and influence of accessible surface area screening." *Macromol Biosci* **10**(7): 746-753.

Brookes, E., B. Demeler, C. Rosano and M. Rocco (2010b). "The implementation of SOMO (SOlution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule." *Eur Biophys J* **39**(3): 423-435.

Brosnan, J. T. and M. E. Brosnan (2006). "Branched-chain amino acids: enzyme and substrate regulation." *J Nutr* **136**(1 Suppl): 207S-211S.

Bryk, R., C. D. Lima, H. Erdjument-Bromage, P. Tempst and C. Nathan (2002). "Metabolic enzymes of mycobacteria linked to antioxidant defense by a thioredoxin-like protein." *Science* **295**(5557): 1073-1077.

Bunik, V. I. (2003). "2-Oxo acid dehydrogenase complexes in redox regulation." *Eur J Biochem* **270**(6): 1036-1042.

Byron, O. (1997). "Construction of hydrodynamic bead models from high-resolution X-ray crystallographic or nuclear magnetic resonance data." *Biophys J* **72**(1): 408-415.

Byron, O. (2008). Hydrodynamic modelling: The solution conformation of macromolecules and their complexes. *Methods in Cell Biology*. J. J. Correia and H. W. Detrich, Elsevier. **84:** 327-373.

Camp, P. J. and D. D. Randall (1985). "Purification and characterization of the pea chloroplast pyruvate dehydrogenase complex : A source of acetyl-CoA and NADH for fatty acid biosynthesis." *Plant Physiol* **77**(3): 571-577.

Chandrasekhar, K., J. Wang, P. Arjunan, M. Sax, Y. H. Park, N. S. Nemeria, S. Kumaran, J. Song, F. Jordan and W. Furey (2013). "Insight to the interaction of the dihydrolipoamide acetyltransferase (E2) core with the peripheral components in the *Escherichia coli* pyruvate dehydrogenase complex via multifaceted structural approaches." *J Biol Chem* **288**(21): 15402-15417.

Chang, H. H., A. M. Falick, P. M. Carlton, J. W. Sedat, J. L. DeRisi and M. A. Marletta (2008). "N-terminal processing of proteins exported by malaria parasites." *Mol Biochem Parasitol* **160**(2): 107-115.

Choudhary, C., C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen and M. Mann (2009). "Lysine acetylation targets protein complexes and co-regulates major cellular functions." *Science* **325**(5942): 834-840.

Choveaux, D. L., J. M. Przyborski and J. P. Goldring (2012). "A *Plasmodium falciparum* copper-binding membrane protein with copper transport motifs." *Malar J* **11**: 397.

Chowdhury, D. R., E. Angov, T. Kariuki and N. Kumar (2009). "A potent malaria transmission blocking vaccine based on codon harmonized full length Pfs48/45 expressed in *Escherichia coli*." *PLoS One* **4**(7): e6352.

Cicek, M., O. Mutlu, A. Erdemir, E. Ozkan, Y. Saricay and D. Turgut-Balik (2013). "Single mutation in Shine-Dalgarno-like sequence present in the amino terminal of lactate dehydrogenase of *Plasmodium* effects the production of an eukaryotic protein expressed in a prokaryotic system." *Mol Biotechnol* **54**(2): 602-608.

Ciszak, E. M., L. G. Korotchkina, P. M. Dominiak, S. Sidhu and M. S. Patel (2003). "Structural basis for flip-flop action of thiamin pyrophosphate-dependent enzymes revealed by human pyruvate dehydrogenase." *J Biol Chem* **278**(23): 21240-21246.

Ciszak, E. M., A. Makal, Y. S. Hong, A. K. Vettaikkorumakankauv, L. G. Korotchkina and M. S. Patel (2006). "How dihydrolipoamide dehydrogenase-binding protein binds dihydrolipoamide dehydrogenase in the human pyruvate dehydrogenase complex." *J Biol Chem* **281**(1): 648-655.

Collins, C. R., S. Das, E. H. Wong, N. Andenmatten, R. Stallmach, F. Hackett, J. P. Herman, S. Müller, M. Meissner and M. J. Blackman (2013). "Robust inducible Cre recombinase activity in the human malaria parasite *Plasmodium falciparum* enables efficient gene deletion within a single asexual erythrocytic growth cycle." *Mol Microbiol* **88**(4): 687-701.

Collins, F. H. and C. A. Hill (2005). *The Anopheles gambiae genome*. Wachington, DC, ASM Press.

Collins, W. E. and G. M. Jeffery (2005). "*Plasmodium ovale*: parasite and disease." *Clin Microbiol Rev* **18**(3): 570-581.

Collins, W. E. and G. M. Jeffery (2007). "*Plasmodium malariae*: parasite and disease." *Clin Microbiol Rev* **20**(4): 579-592.

Conner, M., T. Krell and J. G. Lindsay (1996). "Identification and purification of a distinct dihydrolipoamide dehydrogenase from pea chloroplasts." *Planta* **200**(2): 195-202.

Cowman, A. F. and B. S. Crabb (2006). "Invasion of red blood cells by malaria parasites." *Cell* **124**(4): 755-766.

Cox-Singh, J., J. Hiu, S. B. Lucas, P. C. Divis, M. Zulkarnaen, P. Chandran, K. T. Wong, P. Adem, S. R. Zaki, B. Singh and S. Krishna (2010). "Severe malaria - a case of fatal *Plasmodium knowlesi* infection with post-mortem findings: a case report." *Malar J* **9**: 10.

Cox-Singh, J. (2012). "Zoonotic malaria: *Plasmodium knowlesi*, an emerging pathogen." *Curr Opin Infect Dis* **25**(5): 530-536.

Crabb, B. S. and A. F. Cowman (1996). "Characterization of promoters and stable transfection by homologous and nonhomologous recombination in *Plasmodium falciparum*." *Proc Natl Acad Sci U S A* **93**(14): 7289-7294.

Crabb, B. S., M. Rug, T. W. Gilberger, J. K. Thompson, T. Triglia, A. G. Maier and A. F. Cowman (2004). "Transfection of the human malaria parasite *Plasmodium falciparum*." *Methods Mol Biol* **270**: 263-276.

Crawford, M. J., N. Thomsen-Zieger, M. Ray, J. Schachtner, D. S. Roos and F. Seeber (2006). "*Toxoplasma gondii* scavenges host-derived lipoic acid despite its de novo synthesis in the apicoplast." *EMBO J* **25**(13): 3214-3222.

Danson, M. J., G. Hale, P. Johnson, R. N. Perham, J. Smith and P. Spragg (1979). "Molecular weight and symmetry of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*." *J Mol Biol* **129**(4): 603-617.

Dardel, F., A. L. Davis, E. D. Laue and R. N. Perham (1993). "Three-dimensional structure of the lipoyl domain from *Bacillus stearothermophilus* pyruvate dehydrogenase multienzyme complex." *J Mol Biol* **229**(4): 1037-1048.

Das, S., H. Basu, R. Korde, R. Tewari and S. Sharma (2012). "Arrest of nuclear division in *Plasmodium* through blockage of erythrocyte surface exposed ribosomal protein P2." *PLoS Pathog* **8**(8): e1002858.

Dave, E., J. R. Guest and M. M. Attwood (1995). "Metabolic engineering in *Escherichia coli*: lowering the lipoyl domain content of the pyruvate dehydrogenase complex adversely affects the growth rate and yield." *Microbiology* **141 (Pt 8)**: 1839-1849.

de Kok, A., A. F. Hengeveld, A. Martin and A. H. Westphal (1998). "The pyruvate dehydrogenase multi-enzyme complex from Gram-negative bacteria." *Biochim Biophys Acta* **1385**(2): 353-366.

Deponte, M., H. C. Hoppe, M. C. Lee, A. G. Maier, D. Richard, M. Rug, T. Spielmann and J. M. Przyborski (2012). "Wherever I may roam: protein and membrane trafficking in *P. falciparum*-infected red blood cells." *Mol Biochem Parasitol* **186**(2): 95-116.

Desjardins, R. E., C. J. Canfield, J. D. Haynes and J. D. Chulay (1979). "Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique." *Antimicrob Agents Chemother* **16**(6): 710-718.

Ding, F. and N. V. Dokholyan (2006). "Emergence of protein fold families through rational design." *PLoS Comput Biol* **2**(7): e85.

Dobano, C., M. Sedegah, W. O. Rogers, S. Kumar, H. Zheng, S. L. Hoffman and D. L. Doolan (2009). "*Plasmodium*: mammalian codon optimization of malaria plasmid DNA vaccines enhances antibody responses but not T cell responses nor protective immunity." *Exp Parasitol* **122**(2): 112-123.

Dokholyan, N. V., S. V. Buldyrev, H. E. Stanley and E. I. Shakhnovich (1998). "Discrete molecular dynamics studies of the folding of a protein-like model." *Fold Des* **3**(6): 577-587.

Doniach, S. (2001). "Changes in biomolecular conformation seen by small angle X-ray scattering." *Chem Rev* **101**: 1763-1778.

Dore, E., C. Frontali, T. Forte and S. Fratarcangeli (1983). "Further studies and electron microscopic characterization of *Plasmodium berghei* DNA." *Mol Biochem Parasitol* **8**(4): 339-352.

Dosztanyi, Z., V. Csizmok, P. Tompa and I. Simon (2005). "IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content." *Bioinformatics* **21**(16): 3433-3434.

Evans, L., D. Gowers, K. Firman and J. Youell (2012). "Enhanced purification and characterization of the *Pf*eIF4A (*Pf*H45) helicase from *Plasmodium falciparum* using a codon-optimised clone." *Protein Expr Purif* **85**(1): 1-8.

Falkard, B., T. R. Kumar, L. S. Hecht, K. A. Matthews, P. P. Henrich, S. Gulati, R. E. Lewis, M. J. Manary, E. A. Winzeler, P. Sinnis, S. T. Prigge, V. Heussler, C. Deschermeier and D. Fidock (2013). "A key role for lipoic acid synthesis during *Plasmodium* liver stage development." *Cell Microbiol*.

Faure, M., J. Bourguignon, M. Neuburger, D. MacHerel, L. Sieker, R. Ober, R. Kahn, C. Cohen-Addad and R. Douce (2000). "Interaction between the lipoamide-containing H-protein and the lipoamide dehydrogenase (L-protein) of the glycine decarboxylase multienzyme system 2. Crystal structures of H- and L-proteins." *Eur J Biochem* **267**(10): 2890-2898.

Feagin, J. E., M. J. Gardner, D. H. Williamson and R. J. Wilson (1991). "The putative mitochondrial genome of *Plasmodium falciparum*." *J Protozool* **38**(3): 243-245.

Feagin, J. E. (1994). "The extrachromosomal DNAs of apicomplexan parasites." *Annu Rev Microbiol* **48**: 81-104.

Ferguson, N., T. D. Sharpe, P. J. Schartau, S. Sato, M. D. Allen, C. M. Johnson, T. J. Rutherford and A. R. Fersht (2005). "Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family." *J Mol Biol* **353**(2): 427-446.

Fichera, M. E. and D. S. Roos (1997). "A plastid organelle as a drug target in apicomplexan parasites." *Nature* **390**(6658): 407-409.

Foth, B. J., L. M. Stimmler, E. Handman, B. S. Crabb, A. N. Hodder and G. I. McFadden (2005). "The malaria parasite *Plasmodium falciparum* has only one pyruvate dehydrogenase complex, which is located in the apicoplast." *Mol Microbiol* **55**(1): 39-53.

Frank, R. A., J. V. Pratap, X. Y. Pei, R. N. Perham and B. F. Luisi (2005). "The molecular origins of specificity in the assembly of a multienzyme complex." *Structure* **13**(8): 1119-1130.

Franke, D. and D. Svergun (2009). "DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering." *J Appl Cryst* **42**: 342-346.

Frevert, U., S. Engelmann, S. Zougbede, J. Stange, B. Ng, K. Matuschewski, L. Liebes and H. Yee (2005). "Intravital observation of *Plasmodium berghei* sporozoite infection of the liver." *PLoS Biol* **3**(6): e192.

Gallo, V., E. Schwarzer, S. Rahlfs, R. H. Schirmer, R. van Zwieten, D. Roos, P. Arese and K. Becker (2009). "Inherited glutathione reductase deficiency and *Plasmodium falciparum* malaria--a case study." *PLoS One* **4**(10): e7303.

Gardner, M. J., P. A. Bates, I. T. Ling, D. J. Moore, S. McCready, M. B. Gunasekera, R. J. Wilson and D. H. Williamson (1988). "Mitochondrial DNA of the human malarial parasite *Plasmodium falciparum*." *Mol Biochem Parasitol* **31**(1): 11-17.

Gardner, M. J., D. H. Williamson and R. J. Wilson (1991). "A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts." *Mol Biochem Parasitol* **44**(1): 115-123.

Gardner, M. J., N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M. S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, G. I. McFadden, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser and B. Barrell (2002a). "Genome sequence of the human malaria parasite *Plasmodium falciparum*." *Nature* **419**(6906): 498-511.

Gardner, M. J., S. J. Shallom, J. M. Carlton, S. L. Salzberg, V. Nene, A. Shoaibi, A. Ciecko, J. Lynn, M. Rizzo, B. Weaver, B. Jarrahi, M. Brenner, B. Parvizi, L. Tallon, A. Moazzez, D. Granger, C. Fujii, C. Hansen, J. Pederson, T. Feldblyum, J. Peterson, B. Suh, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, O. White, L. M. Cummings, H. O. Smith, M. D. Adams, J. C. Venter, D. J. Carucci, S. L. Hoffman and C. M. Fraser (2002b). "Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14." *Nature* **419**(6906): 531-534.

Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. Wilkins, R. Appel and A. Bairoch (2005). Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*. J. Walker, Humana Press**:** 571-607.

Gawthorne, J. A., L. E. Reddick, S. N. Akpunarlieva, K. S. Beckham, J. M. Christie, N. M. Alto, M. Gabrielsen and A. J. Roe (2012). "Express your LOV: an engineered flavoprotein as a reporter for protein expression and purification." *PLoS One* **7**(12): e52962.

Gekko, K. (1981). "Enthalpy and entropy of transfer of amino acids and diglycine from water to aqueous polyol solutions." *J Biochem* **90**(6): 1643-1652.

Glushakova, S., G. Humphrey, E. Leikina, A. Balaban, J. Miller and J. Zimmerberg (2010). "New stages in the program of malaria parasite egress imaged in normal and sickle erythrocytes." *Curr Biol* **20**(12): 1117-1121.

Goldberg, D. E. and A. F. Cowman (2010). "Moving in and renovating: exporting proteins from *Plasmodium* into host erythrocytes." *Nat Rev Microbiol* **8**(9): 617-621.

Goodman, C. D. and G. I. McFadden (2013). "Targeting apicoplasts in malaria parasites." *Expert Opin Ther Targets* **17**(2): 167-177.

Goraca, A., H. Huk-Kolega, A. Piechota, P. Kleniewska, E. Ciejka and B. Skibska (2011). "Lipoic acid - biological activity and therapeutic potential." *Pharmacol Rep* **63**(4): 849-858.

Goujon, M., H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern and R. Lopez (2010). "A new bioinformatics analysis tools framework at EMBL-EBI." *Nucleic Acids Res* **38**(Web Server issue): W695-699.

Goyal, M., A. Alam, M. S. Iqbal, S. Dey, S. Bindu, C. Pal, A. Banerjee, S. Chakrabarti and U. Bandyopadhyay (2012). "Identification and molecular characterization of an Alba-family protein from human malaria parasite *Plasmodium falciparum*." *Nucleic Acids Res* **40**(3): 1174-1190.

Griffith, O. W. (1982). "Mechanism of action, metabolism, and toxicity of buthionine sulfoximine and its higher homologs, potent inhibitors of glutathione synthesis." *J Biol Chem* **257**(22): 13704-13712.

Grobusch, M. P. and P. G. Kremsner (2005). "Uncomplicated malaria." *Curr Top Microbiol Immunol* **295**: 83-104.

Gu, Y., Z. H. Zhou, D. B. McCarthy, L. J. Reed and J. K. Stoops (2003). "3D electron microscopy reveals the variable deposition and protein dynamics of the peripheral pyruvate dehydrogenase component about the core." *Proc Natl Acad Sci U S A* **100**(12): 7015-7020.

Guest, J. R., M. M. Attwood, R. S. Machado, K. Y. Matqi, J. E. Shaw and S. L. Turner (1997). "Enzymological and physiological consequences of restructuring the lipoyl domain content of the pyruvate dehydrogenase complex of *Escherichia coli*." *Microbiology* **143 (Pt 2)**: 457-466.

Günther, S., L. Wallace, E. M. Patzewitz, P. J. McMillan, J. Storm, C. Wrenger, R. Bissett, T. K. Smith and S. Müller (2007). "Apicoplast lipoic acid protein ligase B is not essential for *Plasmodium falciparum*." *PLoS Pathog* **3**(12): e189.

Hall, N., A. Pain, M. Berriman, C. Churcher, B. Harris, D. Harris, K. Mungall, S. Bowman, R. Atkin, S. Baker, A. Barron, K. Brooks, C. O. Buckee, C. Burrows, I. Cherevach, C. Chillingworth, T. Chillingworth, Z. Christodoulou, L. Clark, R. Clark, C. Corton, A. Cronin, R. Davies, P. Davis, P. Dear, F. Dearden, J. Doggett, T. Feltwell, A. Goble, I. Goodhead, R. Gwilliam, N. Hamlin, Z. Hance, D. Harper, H. Hauser, T. Hornsby, S. Holroyd, P. Horrocks, S. Humphray, K. Jagels, K. D. James, D. Johnson, A. Kerhornou, A. Knights, B. Konfortov, S. Kyes, N. Larke, D. Lawson, N. Lennard, A. Line, M. Maddison, J. McLean, P. Mooney, S. Moule, L. Murphy, K. Oliver, D. Ormond, C. Price, M. A. Quail, E. Rabbinowitsch, M. A. Rajandream, S. Rutter, K. M. Rutherford, M. Sanders, M. Simmonds, K. Seeger, S. Sharp, R. Smith, R. Squares, S. Squares, K. Stevens, K. Taylor, A. Tivey, L. Unwin, S. Whitehead, J. Woodward, J. E. Sulston, A. Craig, C. Newbold and B. G. Barrell (2002). "Sequence of *Plasmodium falciparum* chromosomes 1, 3-9 and 13." *Nature* **419**(6906): 527-531.

Hanemaaijer, R., A. de Kok, J. Jolles and C. Veeger (1987). "The domain structure of the dihydrolipoyl transacetylase component of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*." *Eur J Biochem* **169**(2): 245-252.

Hanemaaijer, R., A. H. Westphal, A. Berg, W. Van Dongen, A. de Kok and C. Veeger (1989). "The gene encoding dihydrolipoyl transacetylase from *Azotobacter vinelandii*. Expression in *Escherichia coli* and activation and isolation of the protein." *Eur J Biochem* **181**(1): 47-53.

Harris, P. K., S. Yeoh, A. R. Dluzewski, R. A. O'Donnell, C. Withers-Martinez, F. Hackett, L. H. Bannister, G. H. Mitchell and M. J. Blackman (2005). "Molecular identification of a malaria merozoite surface sheddase." *PLoS Pathog* **1**(3): 241-251.

Heath, R. J., J. R. Rubin, D. R. Holland, E. Zhang, M. E. Snow and C. O. Rock (1999). "Mechanism of triclosan inhibition of bacterial fatty acid synthesis." *J Biol Chem* **274**(16): 11110-11114.

Heath, R. J., J. Li, G. E. Roland and C. O. Rock (2000). "Inhibition of the *Staphylococcus aureus* NADPH-dependent enoyl-acyl carrier protein reductase by triclosan and hexachlorophene." *J Biol Chem* **275**(7): 4654-4659.

Heath, C., M. G. Posner, H. C. Aass, A. Upadhyay, D. J. Scott, D. W. Hough and M. J. Danson (2007). "The 2-oxoacid dehydrogenase multi-enzyme complex of the archaeon *Thermoplasma acidophilum* - recombinant expression, assembly and characterization." *FEBS J* **274**(20): 5406-5415.

Hendle, J., A. Mattevi, A.H. Westphal, J. Spee, A. de Kok, A. Teplyakov, W.G. Hol (1995). "Crystallographic and enzymatic investigations on the role of Ser558,

His610, and Asn614 in the catalytic mechanism of *Azotobacter vinelandii* dihydrolipoamide acetyltransferase (E2p)" *Biochemistry* **34**(13): 4287-4298.

Hepler, P. K., C. G. Huff and H. Sprinz (1966). "The fine structure of the exoerythrocytic stages of *Plasmodium fallax*." *J Cell Biol* **30**(2): 333-358.

Hillier, C. J., L. A. Ware, A. Barbosa, E. Angov, J. A. Lyon, D. G. Heppner and D. E. Lanar (2005). "Process development and analysis of liver-stage antigen 1, a preerythrocyte-stage protein-based vaccine for *Plasmodium falciparum*." *Infect Immun* **73**(4): 2109-2115.

Hiromasa, Y., T. Fujisawa, Y. Aso and T. E. Roche (2004). "Organization of the cores of the mammalian pyruvate dehydrogenase complex formed by E2 and E2 plus the E3-binding protein and their capacities to bind the E1 and E3 components." *J Biol Chem* **279**(8): 6921-6933.

Hirschfield, G. M. and M. E. Gershwin (2013). "The immunobiology and pathophysiology of primary biliary cirrhosis." *Annu Rev Pathol* **8**: 303-330.

Holecek, M. (2002). "Relation between glutamine, branched-chain amino acids, and protein metabolism." *Nutrition* **18**(2): 130-133.

Holz, G. G., Jr. (1977). "Lipids and the malarial parasite." *Bull World Health Organ* **55**(2-3): 237-248.

Holz, G. G., Jr., D. H. Beach and I. W. Sherman (1977). "Octadecenoic fatty acids and their association with hemolysis in malaria." *J Protozool* **24**(4): 566-574.

Hopkins, J., R. Fowler, S. Krishna, I. Wilson, G. Mitchell and L. Bannister (1999). "The plastid in *Plasmodium falciparum* asexual blood stages: a three-dimensional ultrastructural analysis." *Protist* **150**(3): 283-295.

Howard, M. J., C. Fuller, R. W. Broadhurst, R. N. Perham, J. G. Tang, J. Quinn, A. G. Diamond and S. J. Yeaman (1998). "Three-dimensional structure of the major autoantigen in primary biliary cirrhosis." *Gastroenterology* **115**(1): 139-146.

Igamberdiev, A. U., N. V. Bykova, W. Ens and R. D. Hill (2004). "Dihydrolipoamide dehydrogenase from porcine heart catalyzes NADH-dependent scavenging of nitric oxide." *FEBS Lett* **568**(1-3): 146-150.

Ikemura, T. (1981). "Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes." *J Mol Biol* **146**(1): 1-21.

Iwanaga, T., T. Sugi, K. Kobayashi, H. Takemae, H. Gong, A. Ishiwa, F. Murakoshi, F. C. Recuenco, T. Horimoto, H. Akashi and K. Kato (2013). "Characterization of *Plasmodium falciparum* cdc2-related kinase and the effects of a CDK inhibitor on the parasites in erythrocytic schizogony." *Parasitol Int* **62**(5): 423-430.

Izard, T., A. Aevarsson, M. D. Allen, A. H. Westphal, R. N. Perham, A. de Kok and W. G. Hol (1999). "Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes." *Proc Natl Acad Sci U S A* **96**(4): 1240-1245.

Jones, D. D., K. M. Stott, M. J. Howard and R. N. Perham (2000). "Restricted motion of the lipoyl-lysine swinging arm in the pyruvate dehydrogenase complex of *Escherichia coli*." *Biochemistry* **39**(29): 8448-8459.

Jones, J. D. and C. D. O'Connor (2011). "Protein acetylation in prokaryotes." *Proteomics* **11**(15): 3012-3022.

Jung, H. I., A. Cooper and R. N. Perham (2003). "Interactions of the peripheral subunit-binding domain of the dihydrolipoyl acetyltransferase component in the assembly of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*." *Eur J Biochem* **270**(22): 4488-4496.

Jung, H. I. and R. N. Perham (2003). "Prediction of the binding site on E1 in the assembly of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*." *FEBS Lett* **555**(2): 405-410.

Jungbauer, A. and W. Kaar (2007). "Current status of technical protein refolding." *J Biotechnol* **128**(3): 587-596.

Kalia, Y. N., S. M. Brocklehurst, D. S. Hipps, E. Appella, K. Sakaguchi and R. N. Perham (1993). "The high-resolution structure of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*." *J Mol Biol* **230**(1): 323-341.

Kane, J. F. (1995). "Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*." *Curr Opin Biotechnol* **6**(5): 494-500.

Kehr, S., N. Sturm, S, Rahlfs, J.M. Przyborski, K. Becker (2010). "Compartmentation of redox metabolism in malaria parasites." *PLoS Pathog* **23**(12): e1001242.

Kelley, L. A., S. P. Gardner and M. J. Sutcliffe (1996). "An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies." *Protein Engineering* **9**: 1063-1065.

Kilejian, A. (1975). "Circular mitochondrial DNA from the avian malarial parasite *Plasmodium lophurae*." *Biochim Biophys Acta* **390**(3): 276-284.

Kim, H., T. C. Liu and M. S. Patel (1991). "Expression of cDNA sequences encoding mature and precursor forms of human dihydrolipoamide dehydrogenase in *Escherichia coli*. Differences in kinetic mechanisms." *J Biol Chem* **266**(15): 9367-9373.

Knapp, J. E., D. T. Mitchell, M. A. Yazdi, S. R. Ernst, L. J. Reed and M. L. Hackert (1998). "Crystal structure of the truncated cubic core component of the *Escherichia coli* 2-oxoglutarate dehydrogenase multienzyme complex." *J Mol Biol* **280**(4): 655-668.

Koch, M. H., P. Vachette and D. I. Svergun (2003). "Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution." *Q Rev Biophys* **36**(2): 147-227.

Komaki-Yasuda, K., S. Kawazu and S. Kano (2003). "Disruption of the *Plasmodium falciparum* 2-Cys peroxiredoxin gene renders parasites hypersensitive to reactive oxygen and nitrogen species." *FEBS Lett* **547**(1-3): 140-144.

Komar, A. A. (2009). "A pause for thought along the co-translational folding pathway." *Trends Biochem Sci* **34**(1): 16-24.

Konarev, P., V. V. Volkov, A. V. Sokolova, M. H. Koch and D. Svergun (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." *J Appl Cryst* **36**: 1277-1282.

Korotchkina, L. G., H. Yang, O. Tirosh, L. Packer and M. S. Patel (2001). "Protection by thiols of the mitochondrial complexes from 4-hydroxy-2-nonenal." *Free Radic Biol Med* **30**(9): 992-999.

Korotchkina, L. G. and M. S. Patel (2008). "Binding of pyruvate dehydrogenase to the core of the human pyruvate dehydrogenase complex." *FEBS Lett* **582**(3): 468-472.

Kozin, M. and D. Svergun (2001). "Automated matching of high- and low-resolution structural methods." *J Appl Cryst* **34**: 33-41.

Kurland, C. and J. Gallant (1996). "Errors of heterologous protein expression." *Curr Opin Biotechnol* **7**(5): 489-493.

LaCount, D. J., L. W. Schoenfeld and S. Fields (2009). "Selection of yeast strains with enhanced expression of *Plasmodium falciparum* proteins." *Mol Biochem Parasitol* **163**(2): 119-122.

Lambros, C. and J. P. Vanderberg (1979). "Synchronization of *Plasmodium falciparum* erythrocytic stages in culture." *J Parasitol* **65**(3): 418-420.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). "Clustal W and Clustal X version 2.0." *Bioinformatics* **23**(21): 2947-2948.

Laue, T. M., B. D. Shah, T. M. Ridgeway and S. L. Pelletier (1992). Computer-aided interpretation of analytical sedimentation data for proteins. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. S. E. Harding, A. J. Rowe and J. C. Horton. London, Royal Society for Chemistry**:** 90-125.

Lebowitz, J., M. S. Lewis and P. Schuck (2002). "Modern analytical ultracentrifugation in protein science: a tutorial review." *Protein Sci* **11**(9): 2067-2079.

Lessard, I. A., G. J. Domingo, A. Borges and R. N. Perham (1998). "Expression of genes encoding the E2 and E3 components of the *Bacillus stearothermophilus* pyruvate dehydrogenase complex and the stoichiometry of subunit interaction in assembly in vitro." *Eur J Biochem* **258**(2): 491-501.

Limenitakis J., D. Soldati-Favre (2011). "Functional genetics in Apicomplexa: potentials and limits." *FEBS Lett* **585**(11): 1579-88.

Lindner, S. E., J. L. Miller and S. H. Kappe (2012). "Malaria parasite pre-erythrocytic infection: preparation meets opportunity." *Cell Microbiol* **14**(3): 316-324.

Lindsay, H., E. Beaumont, S. D. Richards, S. M. Kelly, S. J. Sanderson, N. C. Price and J. G. Lindsay (2000). "FAD insertion is essential for attaining the assembly competence of the dihydrolipoamide dehydrogenase (E3) monomer from *Escherichia coli*." *J Biol Chem* **275**(47): 36665-36670.

Lopez-Barragan, M. J., J. Lemieux, M. Quinones, K. C. Williamson, A. Molina-Cruz, K. Cui, C. Barillas-Mury, K. Zhao and X. Z. Su (2011). "Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*." *BMC Genomics* **12**: 587.

Luersen, K., R. D. Walter and S. Muller (2000). "*Plasmodium falciparum*-infected red blood cells depend on a functional glutathione de novo synthesis attributable to an enhanced loss of glutathione." *Biochem J* **346 Pt 2**: 545-552.

Maier, A. G., J. A. Braks, A. P. Waters and A. F. Cowman (2006). "Negative selection using yeast cytosine deaminase/uracil phosphoribosyl transferase in *Plasmodium falciparum* for targeted gene deletion by double crossover recombination." *Mol Biochem Parasitol* **150**(1): 118-121.

Mande, S. S., S. Sarfaty, M. D. Allen, R. N. Perham and W. G. Hol (1996). "Protein-protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase." *Structure* **4**(3): 277-286.

Marfurt, J., F. Chalfein, P. Prayoga, F. Wabiser, E. Kenangalem, K. A. Piera, D. P. Fairlie, E. Tjitra, N. M. Anstey, K. T. Andrews and R. N. Price (2011). "Ex vivo activity of histone deacetylase inhibitors against multidrug-resistant clinical isolates of *Plasmodium falciparum* and *P. vivax*." *Antimicrob Agents Chemother* **55**(3): 961-966.

Mattevi, A., A. J. Schierbeek and W. G. Hol (1991). "Refined crystal structure of lipoamide dehydrogenase from *Azotobacter vinelandii* at 2.2 Å resolution. A comparison with the structure of glutathione reductase." *J Mol Biol* **220**(4): 975-994.

Mattevi, A., G. Obmolova, E. Schulze, K. H. Kalk, A. H. Westphal, A. de Kok and W. G. Hol (1992a). "Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex." *Science* **255**(5051): 1544-1550.

Mattevi, A., G. Obmolova, J. R. Sokatch, C. Betzel and W. G. Hol (1992b). "The refined crystal structure of *Pseudomonas putida* lipoamide dehydrogenase complexed with NAD+ at 2.45 Å resolution." *Proteins* **13**(4): 336-351.

Mattevi, A., G. Obmolova, K. H. Kalk, A. Teplyakov and W. G. Hol (1993a). "Crystallographic analysis of substrate binding and catalysis in dihydrolipoyl transacetylase (E2p)." *Biochemistry* **32**(15): 3887-3901.

Mattevi, A., G. Obmolova, K. H. Kalk, W. J. van Berkel and W. G. Hol (1993b). "Three-dimensional structure of lipoamide dehydrogenase from *Pseudomonas fluorescens*

at 2.8 Å resolution. Analysis of redox and thermostability properties." *J Mol Biol* **230**(4): 1200-1215.

McCormack, J. G. and R. M. Denton (1979). "The effects of calcium ions and adenine nucleotides on the activity of pig heart 2-oxoglutarate dehydrogenase complex." *Biochem J* **180**(3): 533-544.

McLeod, R., S. P. Muench, J. B. Rafferty, D. E. Kyle, E. J. Mui, M. J. Kirisits, D. G. Mack, C. W. Roberts, B. U. Samuel, R. E. Lyons, M. Dorris, W. K. Milhous and D. W. Rice (2001). "Triclosan inhibits the growth of *Plasmodium falciparum* and *Toxoplasma gondii* by inhibition of apicomplexan Fab I." *Int J Parasitol* **31**(2): 109-113.

McMillan, P. J., L. M. Stimmler, B. J. Foth, G. I. McFadden and S. Müller (2005). "The human malaria parasite *Plasmodium falciparum* possesses two distinct dihydrolipoamide dehydrogenases." *Mol Microbiol* **55**(1): 27-38.

McMillan, P. J. (2006). Thiredoxin reductase and dihydrolipoamide dehydrogenases of *Plasmodium falciparum*. *Division of Infection and Immunity*. Glasgow, University of Glasgow. **Doctor of Philosophy**.

McMurry, L. M., M. Oethinger and S. B. Levy (1998). "Triclosan targets lipid synthesis." *Nature* **394**(6693): 531-532.

Mehlin, C., E. Boni, F. S. Buckner, L. Engel, T. Feist, M. H. Gelb, L. Haji, D. Kim, C. Liu, N. Mueller, P. J. Myler, J. T. Reddy, J. N. Sampson, E. Subramanian, W. C. Van Voorhis, E. Worthey, F. Zucker and W. G. Hol (2006). "Heterologous expression of proteins from *Plasmodium falciparum*: results from 1000 genes." *Mol Biochem Parasitol* **148**(2): 144-160.

Meierjohann, S., R. D. Walter and S. Müller (2002). "Regulation of intracellular glutathione levels in erythrocytes infected with chloroquine-sensitive and chloroquine-resistant *Plasmodium falciparum*." *Biochem J* **368**(Pt 3): 761-768.

Mertens, H. D. and D. I. Svergun (2010). "Structural characterization of proteins and complexes using small-angle X-ray solution scattering." *J Struct Biol* **172**(1): 128-141.

Miao, J., Q. Fan, L. Cui and J. Li (2006). "The malaria parasite *Plasmodium falciparum* histones: organization, expression, and acetylation." *Gene* **369**: 53-65.

Miao, J., Q. Fan, L. Cui, X. Li, H. Wang, G. Ning and J. C. Reese (2010). "The MYST family histone acetyltransferase regulates gene expression and cell cycle in malaria parasite *Plasmodium falciparum*." *Mol Microbiol* **78**(4): 883-902.

Miao, J., M. Lawrence, V. Jeffers, F. Zhao, D. Parker, Y. Ge, W. J. Sullivan, Jr. and L. Cui (2013). "Extensive lysine acetylation occurs in evolutionarily conserved metabolic pathways and parasite-specific functions during *Plasmodium falciparum* intraerythrocytic development." *Mol Microbiol* **89**(4): 660-675.

Mikolajczak, S. A. and S. H. Kappe (2006). "A clash to conquer: the malaria parasite liver infection." *Mol Microbiol* **62**(6): 1499-1506.

Millar, A. H., C. J. Leaver and S. A. Hill (1999). "Characterization of the dihydrolipoamide acetyltransferase of the mitochondrial pyruvate dehydrogenase complex from potato and comparisons with similar enzymes in diverse plant species." *Eur J Biochem* **264**(3): 973-981.

Milne, J. L., D. Shi, P. B. Rosenthal, J. S. Sunshine, G. J. Domingo, X. Wu, B. R. Brooks, R. N. Perham, R. Henderson and S. Subramaniam (2002). "Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine." *EMBO J* **21**(21): 5587-5598.

Milne, J. L., X. Wu, M. J. Borgnia, J. S. Lengyel, B. R. Brooks, D. Shi, R. N. Perham and S. Subramaniam (2006). "Molecular structure of a 9 MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy." *J Biol Chem* **281**(7): 4364-4370.

Mooney, B. P., J. A. Miernyk and D. D. Randall (1999). "Cloning and characterization of the dihydrolipoamide S-acetyltransferase subunit of the plastid pyruvate dehydrogenase complex (E2) from *Arabidopsis*." *Plant Physiol* **120**(2): 443-452.

Mooney, B. P., J. A. Miernyk and D. D. Randall (2002). "The complex fate of alpha-ketoacids." *Annu Rev Plant Biol* **53**: 357-375.

Mota, M. M., G. Pradel, J. P. Vanderberg, J. C. Hafalla, U. Frevert, R. S. Nussenzweig, V. Nussenzweig and A. Rodriguez (2001). "Migration of *Plasmodium* sporozoites through cells before infection." *Science* **291**(5501): 141-144.

Mueller, I., M. R. Galinski, J. K. Baird, J. M. Carlton, D. K. Kochar, P. L. Alonso and H. A. del Portillo (2009). "Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite." *Lancet Infect Dis* **9**(9): 555-566.

Nakajima, C., H. Kuyama, T. Nakazawa, O. Nishimura (2012a). "C-terminal sequencing of protein by MALDI mass spectrometry through the specific derivatization of the α-carboxyl group with 3-aminopropyltris-(2,4,6-trimethoxyphenyl)phosphonium bromide." *Anal Bioanal Chem* **404**(1):125-32.

Nakajima, C., H. Kuyama, K. Tanaka (2012b) "Mass spectrometry-based sequencing of protein C-terminal peptide using α-carboxyl group-specific derivatization and COOH capturing." *Anal Biochem* **428**(2):167-72.

Neagle, J., O. De Marcucci, B. Dunbar and J. G. Lindsay (1989). "Component X of mammalian pyruvate dehydrogenase complex: structural and functional relationship to the lipoate acetyltransferase (E2) component." *FEBS Lett* **253**(1-2): 11-15.

Neuweiler, H., T. D. Sharpe, T. J. Rutherford, C. M. Johnson, M. D. Allen, N. Ferguson and A. R. Fersht (2009). "The folding mechanism of BBL: Plasticity of transition-state structure observed within an ultrafast folding protein family." *J Mol Biol* **390**(5): 1060-1073.

Norris, K. L., J. Y. Lee and T. P. Yao (2009). "Acetylation goes global: the emergence of acetylation biology." *Sci Signal* **2**(97): pe76.

O'Connor, T. P., T. E. Roche and J. V. Paukstelis (1982). "13C nuclear magnetic resonance study of the pyruvate dehydrogenase-catalyzed acetylation of dihydrolipoamide." *J Biol Chem* **257**(6): 3110-3112.

O'Donnell, R. A., L. H. Freitas-Junior, P. R. Preiser, D. H. Williamson, M. Duraisingh, T. F. McElwain, A. Scherf, A. F. Cowman and B. S. Crabb (2002). "A genetic screen for improved plasmid segregation reveals a role for Rep20 in the interaction of *Plasmodium falciparum* chromosomes." *EMBO J* **21**(5): 1231-1239.

Odessey, R. (1982). "Purification of rat kidney branched-chain oxo acid dehydrogenase complex with endogenous kinase activity." *Biochem J* **204**(1): 353-356.

Okada, K. (2009). "The novel heme oxygenase-like protein from *Plasmodium falciparum* converts heme to bilirubin IXalpha in the apicoplast." *FEBS Lett* **583**(2): 313-319.

Olinares, P. D., L. Ponnala and K. J. van Wijk (2010). "Megadalton complexes in the chloroplast stroma of *Arabidopsis thaliana* characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering." *Mol Cell Proteomics* **9**(7): 1594-1615.

Olsson, J. M., L. Xia, L. C. Eriksson and M. Bjornstedt (1999). "Ubiquinone is reduced by lipoamide dehydrogenase and this reaction is potently stimulated by zinc." *FEBS Lett* **448**(1): 190-192.

Packard, R. M. (2007). *The making of a tropical disease: A short history of malaria* Baltimore, The Johns Hopkins University Press

Packman, L. C., R. N. Perham and G. C. Roberts (1984). "Domain structure and 1H NMR spectroscopy of the pyruvate dehydrogenase complex of *Bacillus stearothermophilus*." *Biochem J* **217**(1): 219-227.

Patel, K. P., T. W. O'Brien, S. H. Subramony, J. Shuster and P. W. Stacpoole (2012). "The spectrum of pyruvate dehydrogenase complex deficiency: clinical, biochemical and genetic features in 371 patients." *Mol Genet Metab* **105**(1): 34-43.

Patel, M. S., N. N. Vettakkorumakankav and T. C. Liu (1995). "Dihydrolipoamide dehydrogenase: activity assays." *Methods Enzymol* **252**: 186-195.

Patel, M. S. and L. G. Korotchkina (2001). "Regulation of mammalian pyruvate dehydrogenase complex by phosphorylation: complexity of multiple phosphorylation sites and kinases." *Exp Mol Med* **33**(4): 191-197.

Patzewitz, E. M., E. H. Wong and S. Müller (2012). "Dissecting the role of glutathione biosynthesis in *Plasmodium falciparum*." *Mol Microbiol* **83**(2): 304-318.

Paul, K. S., C. J. Bacchi and P. T. Englund (2004). "Multiple triclosan targets in *Trypanosoma brucei*." *Eukaryot Cell* **3**(4): 855-861.

Pei, X. Y., C. M. Titman, R. A. Frank, F. J. Leeper and B. F. Luisi (2008). "Snapshots of catalysis in the E1 subunit of the pyruvate dehydrogenase multienzyme complex." *Structure* **16**(12): 1860-1872.

Pei, Y., A. S. Tarun, A. M. Vaughan, R. W. Herman, J. M. Soliman, A. Erickson-Wayman and S. H. Kappe (2010). "*Plasmodium* pyruvate dehydrogenase activity is only essential for the parasite's progression from liver infection to blood infection." *Mol Microbiol* **75**(4): 957-971.

Perham, R. N., B. Leistler, R. G. Solomon and P. Guptasarma (1996). "Protein engineering of domains in flavoprotein disulphide oxidoreductases: contributions to folding and assembly." *Biochem Soc Trans* **24**(1): 61-66.

Perham, R. N. and P. A. Reche (1998). "Swinging arms in multifunctional enzymes and the specificity of post-translational modification." *Biochem Soc Trans* **26**(3): 299-303.

Perozzo, R., M. Kuo, A. Sidhu, J. T. Valiyaveettil, R. Bittman, W. R. Jacobs, Jr., D. A. Fidock and J. C. Sacchettini (2002). "Structural elucidation of the specificity of the antibacterial agent triclosan for malarial enoyl acyl carrier protein reductase." *J Biol Chem* **277**(15): 13106-13114.

Petoukhov, M., P. Konarev, A. Kikhney and D. Svergun (2007). "ATSAS 2.1-towards automated and web-supported small-angle data analysis." *J Appl Cryst* **40**: s223-s228.

Petoukhov, M., D. Franke, A. V. Shkumatov, G. Tria, A. Kikhney, M. Gadja, C. Gorba, H. D. Mertens, P. Konarev and D. Svergun (2012). "New developments in the ATSAS program package for small-angle scattering data analysis." *J Appl Cryst* **45**: 342-350.

Petsko, G. A. (1991). "Enzyme evolution. Deja vu all over again." *Nature* **352**(6331): 104-105.

Pettit, F. H., S. J. Yeaman and L. J. Reed (1978). "Purification and characterization of branched chain alpha-keto acid dehydrogenase complex of bovine kidney." *Proc Natl Acad Sci U S A* **75**(10): 4881-4885.

Pizzi, E. and C. Frontali (2001). "Low-complexity regions in *Plasmodium falciparum* proteins." *Genome Res* **11**(2): 218-229.

Qidwai, T. and F. Khan (2012). "Antimalarial drugs and drug targets specific to fatty acid metabolic pathway of *Plasmodium falciparum*." *Chem Biol Drug Des* **80**(2): 155-172.

Rai, N., M. Nollmann, B. Spotorno, G. Tassara, O. Byron and M. Rocco (2005). "SOMO (SOlution MOdeler) differences between X-Ray- and NMR-derived bead models suggest a role for side chain flexibility in protein hydrodynamics." *Structure* **13**(5): 723-734.

Ralph, S. A., G. G. van Dooren, R. F. Waller, M. J. Crawford, M. J. Fraunholz, B. J. Foth, C. J. Tonkin, D. S. Roos and G. I. McFadden (2004). "Tropical infectious diseases:

metabolic maps and functions of the *Plasmodium falciparum* apicoplast." *Nat Rev Microbiol* **2**(3): 203-216.

Ralston, G. (1993). *Introduction to Analytical Ultracentrifugation*. Fullerton, USA, Beckman.

Reed, L. J. (2001). "A trail of research from lipoic acid to alpha-keto acid dehydrogenase complexes." *J Biol Chem* **276**(42): 38329-38336.

Rock, R. C. (1971). "Incorporation of 14 C-labelled non-lipid precursors into lipid of *Plasmodium knowlesi* in vitro." *Comp Biochem Physiol B* **40**(3): 657-669.

Rock, R. C., J. Standefer and W. Little (1971). "Incorporation of 33p-orthophosphate into membrane phospholipids of *Plasmodium knowlesi* and host erythrocytes of *Macaca mullatta*." *Comp Biochem Physiol B* **40**(2): 543-561.

Roessle, M., R. Klaering, U. Ristau, B. Robrahn, D. Jahn, T. Gehrmann, P. Konarev, A. Round, S. Fiedler, C. Hermes and D. Svergun (2007). "Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg." *J Appl Cryst* **40**: s190-s194.

Rosenfeld, L. (1982). Svedberg and the Ultracentrifuge. *Origins of Clinical Chemistry: The Evolution of Protein Analysis* New York, Academic Press**:** 144-161.

Roy, A., A. Kucukural and Y. Zhang (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." *Nat Protoc* **5**(4): 725-738.

Russell, G. C. and J. R. Guest (1991). "Sequence similarities within the family of dihydrolipoamide acyltransferases and discovery of a previously unidentified fungal enzyme." *Biochim Biophys Acta* **1076**(2): 225-232.

Sanders, P. R., L. M. Kats, D. R. Drew, R. A. O'Donnell, M. O'Neill, A. G. Maier, R. L. Coppel and B. S. Crabb (2006). "A set of glycosylphosphatidyl inositol-anchored membrane proteins of *Plasmodium falciparum* is refractory to genetic deletion." *Infect Immun* **74**(7): 4330-4338.

Sato, S., B. Clough, L. Coates and R. J. Wilson (2004). "Enzymes for heme biosynthesis are found in both the mitochondrion and plastid of the malaria parasite *Plasmodium falciparum*." *Protist* **155**(1): 117-125.

Saul, A. and D. Battistutta (1988). "Codon usage in *Plasmodium falciparum*." *Mol Biochem Parasitol* **27**: 35-42.

Schirf, V. and K. L. Planken (2008). Analytical Ultracentrifuge User Guide Volume 1: Hardware. San Antonio, UT Health Science Center.

Schmitz, S., M. Grainger, S. Howell, L. J. Calder, M. Gaeb, J. C. Pinder, A. A. Holder and C. Veigel (2005). "Malaria parasite actin filaments are very short." *J Mol Biol* **349**(1): 113-125.

Schöneck, R., O. Billaut-Mulot, P. Numrich, M. A. Ouaissi and R. L. Krauth-Siegel (1997). "Cloning, sequencing and functional expression of dihydrolipoamide dehydrogenase from the human pathogen *Trypanosoma cruzi*." *Eur J Biochem* **243**(3): 739-747.

Schuck, P. (2000). "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling." *Biophys J* **78**(3): 1606-1619.

Schuck, P., M. A. Perugini, N. R. Gonzales, G. J. Howlett and D. Schubert (2002). "Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems." *Biophys J* **82**(2): 1096-1111.

Schuck, P. (2003). "On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation." *Anal Biochem* **320**(1): 104-124.

Schulze, E., A. H. Westphal, H. Boumans and A. de Kok (1991a). "Site-directed mutagenesis of the dihydrolipoyl transacetylase component (E2p) of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Binding of the peripheral components E1p and E3." *Eur J Biochem* **202**(3): 841-848.

Schulze, E., A. H. Westphal, G. Obmolova, A. Mattevi, W. G. Hol and A. de Kok (1991b). "The catalytic domain of the dihydrolipoyl transacetylase component of the pyruvate dehydrogenase complex from *Azotobacter vinelandii* and *Escherichia coli*. Expression, purification, properties and preliminary X-ray analysis." *Eur J Biochem* **201**(3): 561-568.

Schulze, E., A. H. Westphal, R. Hanemaaijer and A. de Kok (1993). "Structure/function relationships in the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Role of the linker region between the binding and catalytic domain of the dihydrolipoyl transacetylase component." *Eur J Biochem* **211**(3): 591-599.

Schuster, T. M. and J. M. Toedt (1996). "New revolutions in the evolution of analytical ultracentrifugation." *Curr Opin Struct Biol* **6**(5): 650-658.

Seeber, F., J. Limenitakis and D. Soldati-Favre (2008). "Apicomplexan mitochondrial metabolism: a story of gains, losses and retentions." *Trends Parasitol* **24**(10): 468-478.

Serdyuk, I. N., N. R. Zaccai and J. Zaccai (2007a). Analytical ultracentrifugation. *Methods in Molecular Biophysics*. Cambridge, Cambridge University Press**:** 339-387.

Serdyuk, I. N., N. R. Zaccai and J. Zaccai (2007b). The macromolecule as a radiation scattering particle. *Methods in Molecular Biophysics*. Cambridge, Cambridge University Press**:** 794-837.

Serdyuk, I. N., N. R. Zaccai and J. Zaccai (2007c). Small-angle scattering. *Methods in Molecular Biophysics*. Cambridge, Cambridge University Press**:** 794-837.

Sherman, I. W. (2005). The life of Plasmodium: an overview. *Molecular Approaches to Malaria.* Washington, DC, ASM Press**:** 3-10

Sherman, I. W. (2009a). The ring road to the apicoplast. *Advances in parasitology: Reflections on a century of malaria biochemistry*. D. Rollinson, Hay, S. London, Academic Press, Elsevier. **67:** 105-110.

Sherman, I. W. (2009b). *In vivo* and *in vitro* models. *Advances in parasitology: Reflections on a century of malaria biochemistry*. D. Rollinson, Hay, S. London, Academic Press, Elsevier. **67:** 27-47.

Smith, C. M., J. Bryla and J. R. Williamson (1974). "Regulation of mitochondrial alpha-ketoglutarate metabolism by product inhibition at alpha-ketoglutarate dehydrogenase." *J Biol Chem* **249**(5): 1497-1505.

Smolle, M., A. E. Prior, A. E. Brown, A. Cooper, O. Byron and J. G. Lindsay (2006). "A new level of architectural complexity in the human pyruvate dehydrogenase complex." *J Biol Chem* **281**(28): 19772-19780.

Starkov, A. A. (2013). "An update on the role of mitochondrial alpha-ketoglutarate dehydrogenase in oxidative stress." *Mol Cell Neurosci* **55**: 13-16.

Stephens, L. L., A. Shonhai and G. L. Blatch (2011). "Co-expression of the *Plasmodium falciparum* molecular chaperone, *Pf*Hsp70, improves the heterologous production of the antimalarial drug target GTP cyclohydrolase I, *Pf*GCHI." *Protein Expr Purif* **77**(2): 159-165.

Stoops, J. K., R. H. Cheng, M. A. Yazdi, C. Y. Maeng, J. P. Schroeter, U. Klueppelberg, S. J. Kolodziej, T. S. Baker and L. J. Reed (1997). "On the unique structural organization of the *Saccharomyces cerevisiae* pyruvate dehydrogenase complex." *J Biol Chem* **272**(9): 5757-5764.

Storm, J. and S. Müller (2012). "Lipoic acid metabolism of *Plasmodium*--a suitable drug target." *Curr Pharm Des* **18**(24): 3480-3489.

Sturm, A., R. Amino, C. van de Sand, T. Regen, S. Retzlaff, A. Rennenberg, A. Krueger, J. M. Pollok, R. Menard and V. T. Heussler (2006). "Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids." *Science* **313**(5791): 1287-1290.

Sturm, A., S. Graewe, B. Franke-Fayard, S. Retzlaff, S. Bolte, B. Roppenser, M. Aepfelbacher, C. Janse and V. Heussler (2009). "Alteration of the parasite plasma

membrane and the parasitophorous vacuole membrane during exo-erythrocytic development of malaria parasites." *Protist* **160**(1): 51-63.

Sullivan, M., J. Li, S. Kumar, M. J. Rogers and T. F. McCutchan (2000). "Effects of interruption of apicoplast function on malaria infection, development, and transmission." *Mol Biochem Parasitol* **109**(1): 17-23.

Suntres, Z. E. (2002). "Role of antioxidants in paraquat toxicity." *Toxicology* **180**(1): 65-77.

Surolia, N. and A. Surolia (2001). "Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum*." *Nat Med* **7**(2): 167-173.

Svedberg, T. and R. Fahraeus (1926). "A new method for the determination of the molecular weight of the proteins." *J Am Chem Soc* **48**: 430-438.

Svedberg, T. and J. B. Nichols (1927). "The application of the oil turbine type of ultracentrifuge to the study of the stability region of carbon monoxide-hemoglobin." *J Am Chem Soc* **49**: 2920-2934.

Svergun, D. (2007). "Small-angle scattering studies of macromolecular solutions." *J Appl Cryst* **40**: s10-s17.

Svergun, D. I. (1992). "Determination of the regularization parameter in indirect-transform methods using perceptual criteria " *J Appl Cryst* **25**: 495-503.

Svergun, D. I. (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing." *Biophys J* **76**(6): 2879-2886.

Thanaraj, T. A. and P. Argos (1996). "Protein secondary structural types are differentially coded on messenger RNA." *Protein Sci* **5**(10): 1973-1983.

Thelen, J. J., M. G. Muszynski, N. R. David, M. H. Luethy, T. E. Elthon, J. A. Miernyk and D. D. Randall (1999). "The dihydrolipoamide S-acetyltransferase subunit of the mitochondrial pyruvate dehydrogenase complex from maize contains a single lipoyl domain." *J Biol Chem* **274**(31): 21769-21775.

Toyoda, T., K. Suzuki, T. Sekiguchi, L. J. Reed and A. Takenaka (1998). "Crystal structure of eukaryotic E3, lipoamide dehydrogenase from yeast." *J Biochem* **123**(4): 668-674.

Trager, W. and J. B. Jensen (1976). "Human malaria parasites in continuous culture." *Science* **193**(4254): 673-675.

Tretter, L. and V. Adam-Vizi (2005). "Alpha-ketoglutarate dehydrogenase: a target and generator of oxidative stress." *Philos Trans R Soc Lond B Biol Sci* **360**(1464): 2335-2345.

Tsunoda, Y., N. Sakai, K. Kikuchi, S. Katoh, K. Akagi, J. Miura-Ohnuma, Y. Tashiro, K. Murata, N. Shibuya and E. Katoh (2005). "Improving expression and solubility of rice proteins produced as fusion proteins in *Escherichia coli*." *Protein Expr Purif* **42**(2): 268-277.

Tuteja, R. (2007). "Malaria - an overview." *FEBS J* **274**(18): 4670-4679.

Vaidya, A. B. and P. Arasu (1987). "Tandemly arranged gene clusters of malarial parasites that are highly conserved and transcribed." *Mol Biochem Parasitol* **22**(2-3): 249-257.

Vaidya, A. B., R. Akella and K. Suplick (1989). "Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite." *Mol Biochem Parasitol* **35**(2): 97-107.

van Dooren, G. G., L. M. Stimmler and G. I. McFadden (2006). "Metabolic maps and functions of the *Plasmodium* mitochondrion." *FEMS Microbiol Rev* **30**(4): 596-630.

van Dooren, G. G. and B. Striepen (2013). "The algal past and parasite present of the apicoplast." *Annu Rev Microbiol* **67**: 271-289.

van Holde, K. E. and J. C. Hansen (1998). "Analytical ultracentrifugation from 1924 to the present: A remarkable history." *Chemtracts: Biochem Mol Biol* **11**: 933-943.

Vanderberg, J. P. and U. Frevert (2004). "Intravital microscopy demonstrating antibody-mediated immobilisation of *Plasmodium berghei* sporozoites injected into skin by mosquitoes." *Int J Parasitol* **34**(9): 991-996.

Vaughan, A. M., M. T. O'Neill, A. S. Tarun, N. Camargo, T. M. Phuong, A. S. Aly, A. F. Cowman and S. H. Kappe (2009). "Type II fatty acid synthesis is essential only for malaria parasite late liver stage development." *Cell Microbiol* **11**(3): 506-520.

Vedadi, M., J. Lew, J. Artz, M. Amani, Y. Zhao, A. Dong, G. A. Wasney, M. Gao, T. Hills, S. Brokx, W. Qiu, S. Sharma, A. Diassiti, Z. Alam, M. Melone, A. Mulichak, A. Wernimont, J. Bray, P. Loppnau, O. Plotnikova, K. Newberry, E. Sundararajan, S. Houston, J. Walker, W. Tempel, A. Bochkarev, I. Kozieradzki, A. Edwards, C. Arrowsmith, D. Roos, K. Kain and R. Hui (2007). "Genome-scale protein expression and structural biology of *Plasmodium falciparum* and related Apicomplexan organisms." *Mol Biochem Parasitol* **151**(1): 100-110.

Vial, H. J. and M. L. Ancelin (1992). "Malarial lipids. An overview." *Subcell Biochem* **18**: 259-306.

Vijayakrishnan, S., S. M. Kelly, R. J. Gilbert, P. Callow, D. Bhella, T. Forsyth, J. G. Lindsay and O. Byron (2010). "Solution structure and characterisation of the human pyruvate dehydrogenase complex core assembly." *J Mol Biol* **399**(1): 71-93.

Vijayakrishnan, S., P. Callow, M. A. Nutley, D. P. McGow, D. Gilbert, P. Kropholler, A. Cooper, O. Byron and J. G. Lindsay (2011). "Variation in the organization and subunit composition of the mammalian pyruvate dehydrogenase complex E2/E3BP core assembly." *Biochem J* **437**(3): 565-574.

Vistica, J., J. Dam, A. Balbo, E. Yikilmaz, R. A. Mariuzza, T. A. Rouault and P. Schuck (2004). "Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition." *Anal Biochem* **326**(2): 234-256.

Volkov, V. V. and D. Svergun (2003). "Uniqueness of *ab initio* shape determination in small-angle scattering." *Journal of Applied Crystallography* **36**: 860-864.

Waller, R. F., P. J. Keeling, R. G. Donald, B. Striepen, E. Handman, N. Lang-Unnasch, A. F. Cowman, G. S. Besra, D. S. Roos and G. I. McFadden (1998). "Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*." *Proc Natl Acad Sci U S A* **95**(21): 12352-12357.

Waller, R. F., M. B. Reed, A. F. Cowman and G. I. McFadden (2000). "Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway." *EMBO J* **19**(8): 1794-1802.

Waller, R. F., S. A. Ralph, M. B. Reed, V. Su, J. D. Douglas, D. E. Minnikin, A. F. Cowman, G. S. Besra and G. I. McFadden (2003). "A type II pathway for fatty acid biosynthesis presents drug targets in *Plasmodium falciparum*." *Antimicrob Agents Chemother* **47**(1): 297-301.

Wallis, N. G. and R. N. Perham (1994). "Structural dependence of post-translational modification and reductive acetylation of the lipoyl domain of the pyruvate dehydrogenase multienzyme complex." *J Mol Biol* **236**(1): 209-216.

Ward, D. E., R. P. Ross, C. C. van der Weijden, J. L. Snoep and A. Claiborne (1999). "Catabolism of branched-chain alpha-keto acids in *Enterococcus faecalis*: the bkd gene cluster, enzymes, and metabolic route." *J Bacteriol* **181**(17): 5433-5442.

Waterhouse, A. M., J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton (2009). "Jalview Version 2-a multiple sequence alignment editor and analysis workbench." *Bioinformatics* **25**: 1189-1191.

Wheatley, N. C., K. T. Andrews, T. L. Tran, A. J. Lucke, R. C. Reid and D. P. Fairlie (2010). "Antimalarial histone deacetylase inhibitors containing cinnamate or NSAID components." *Bioorg Med Chem Lett* **20**(23): 7080-7084.

White, N. J. (2008). "The role of anti-malarial drugs in eliminating malaria." *Malar J* **7 Suppl 1**: S8.

WHO (2005). World Malaria Report.
WHO (2009). World Malaria Report.
WHO (2010). Global report on antimalarial drug efficacy and drug resistance.
WHO (2012). World Malaria Report.
Williamson, D. H., R. J. Wilson, P. A. Bates, S. McCready, F. Perler and B. U. Qiang (1985). "Nuclear and mitochondrial DNA of the primate malarial parasite *Plasmodium knowlesi*." *Mol Biochem Parasitol* **14**(2): 199-209.
Williamson, D. H., M. J. Gardner, P. Preiser, D. J. Moore, K. Rangachari and R. J. Wilson (1994). "The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry." *Mol Gen Genet* **243**(2): 249-252.
Willms, C. R., R. M. Oliver, H. R. Henney, Jr., B. B. Mukherjee and L. J. Reed (1967). "Alpha-keto acid dehydrogenase complexes. VI. Dissociation and reconstitution of the dihydrolipoyl transacetylase of *Escherichia coli*." *J Biol Chem* **242**(5): 889-897.
Wilson, R. J., M. J. Gardner, J. E. Feagin and D. H. Williamson (1991). "Have malaria parasites three genomes?" *Parasitol Today* **7**(6): 134-136.
Wilson, R. J., M. Fry, M. J. Gardner, J. E. Feagin and D. H. Williamson (1992). "Subcellular fractionation of the two organelle DNAs of malaria parasites." *Curr Genet* **21**(4-5): 405-408.
Wilson, R. J., P. W. Denny, P. R. Preiser, K. Rangachari, K. Roberts, A. Roy, A. Whyte, M. Strath, D. J. Moore, P. W. Moore and D. H. Williamson (1996). "Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*." *J Mol Biol* **261**(2): 155-172.
Wirth, C. C. and G. Pradel (2012). "Molecular mechanisms of host cell egress by malaria parasites." *Int J Med Microbiol* **302**(4-5): 172-178.
Wirth, D. F. (2002). "Biological revelations." *Nature* **419**(6906): 495-496.
Withers-Martinez, C., C. Suarez, S. Fulle, S. Kher, M. Penzo, J. P. Ebejer, K. Koussis, F. Hackett, A. Jirgensons, P. Finn and M. J. Blackman (2012). "*Plasmodium* subtilisin-like protease 1 (SUB1): insights into the active-site structure, specificity and function of a pan-malaria drug target." *Int J Parasitol* **42**(6): 597-612.
Wrenger, C. and S. Müller (2004). "The human malaria parasite *Plasmodium falciparum* has distinct organelle-specific lipoylation pathways." *Mol Microbiol* **53**(1): 103-113.
Xia, L., M. Bjornstedt, T. Nordman, L. C. Eriksson and J. M. Olsson (2001). "Reduction of ubiquinone by lipoamide dehydrogenase. An antioxidant regenerating pathway." *Eur J Biochem* **268**(5): 1486-1490.
Xiao, H., T. Tanaka, M. Ogawa and R. Y. Yada (2007). "Expression and enzymatic characterization of the soluble recombinant plasmepsin I from *Plasmodium falciparum*." *Protein Eng Des Sel* **20**(12): 625-633.
Xing, S. and Y. Poirier (2012). "The protein acetylome and the regulation of metabolism." *Trends Plant Sci* **17**(7): 423-430.
Yang, D., J. Song, T. Wagenknecht and T. E. Roche (1997). "Assembly and full functionality of recombinantly expressed dihydrolipoyl acetyltransferase component of the human pyruvate dehydrogenase complex." *J Biol Chem* **272**(10): 6361-6369.
Yang, Y. S. and P. A. Frey (1986). "Dihydrolipoyl transacetylase of *Escherichia coli*. Formation of 8-S-acetyldihydrolipoamide." *Biochemistry* **25**(25): 8173-8178.
Yeoh, S., R. A. O'Donnell, K. Koussis, A. R. Dluzewski, K. H. Ansell, S. A. Osborne, F. Hackett, C. Withers-Martinez, G. H. Mitchell, L. H. Bannister, J. S. Bryans, C. A. Kettleborough and M. J. Blackman (2007). "Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes." *Cell* **131**(6): 1072-1083.

Yu, M., T. R. Kumar, L. J. Nkrumah, A. Coppi, S. Retzlaff, C. D. Li, B. J. Kelly, P. A. Moura, V. Lakshmanan, J. S. Freundlich, J. C. Valderramos, C. Vilcheze, M. Siedner, J. H. Tsai, B. Falkard, A. B. Sidhu, L. A. Purcell, P. Gratraud, L. Kremer, A. P. Waters, G. Schiehser, D. P. Jacobus, C. J. Janse, A. Ager, W. R. Jacobs, Jr., J. C. Sacchettini, V. Heussler, P. Sinnis and D. A. Fidock (2008a). "The fatty acid biosynthesis enzyme FabI plays a key role in the development of liver-stage malarial parasites." *Cell Host Microbe* **4**(6): 567-578.

Yu, X., Y. Hiromasa, H. Tsen, J. K. Stoops, T. E. Roche and Z. H. Zhou (2008b). "Structures of the human pyruvate dehydrogenase complex cores: a highly conserved catalytic center with flexible N-terminal domains." *Structure* **16**(1): 104-114.

Zhang, Y. (2008). "I-TASSER server for protein 3D structure prediction." *BMC Bioinformatics* **9**: 40.

Zhou, Z., P. Schnake, L. Xiao and A. A. Lal (2004). "Enhanced expression of a recombinant malaria candidate vaccine in *Escherichia coli* by codon optimization." *Protein Expr Purif* **34**(1): 87-94.

Zhou, Z. H., W. Liao, R. H. Cheng, J. E. Lawson, D. B. McCarthy, L. J. Reed and J. K. Stoops (2001a). "Direct evidence for the size and conformational variability of the pyruvate dehydrogenase complex revealed by three-dimensional electron microscopy. The "breathing" core and its functional relationship to protein dynamics." *J Biol Chem* **276**(24): 21704-21713.

Zhou, Z. H., D. B. McCarthy, C. M. O'Connor, L. J. Reed and J. K. Stoops (2001b). "The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes." *Proc Natl Acad Sci U S A* **98**(26): 14802-14807.